THE UNIVERSITY
of EDINBURGH

Detection and Characterisation of Genetic Associations with Canine Skull Shape and Disease

Thomas W. Marchant

Doctor of Philosophy

College of Medicine and Veterinary Medicine, The University of Edinburgh

2018

Declaration

I declare that the work contained within this thesis is my own with the exception of specific experiments which are clearly indicated in text. The thesis has been composed by myself and has not been submitted for any other degree or qualification.

Thomas W. Marchant

Lay Summary

Changes in the size and shape of the head are common across many different animals and are perhaps most noticeable in the dog. Popular dog breeds such as the Greyhound have long, narrow heads whilst breeds such as the Pug have very flat faces. Dog breeders have slowly introduced the flat-faced look into many of these breeds by carefully controlling which dogs have litters. An unintentional side effect of these breeding programmes has been the introduction of many diseases into the flat-faced dogs. One of the most common diseases are a group of breathing problems called Brachycephalic Obstructive Airway Syndrome (BOAS). Dogs that suffer from BOAS have difficulty breathing and often need to have surgery to improve their symptoms. This is a major welfare concern.

The objective of this project was to improve the welfare of dogs by finding the DNA changes that cause differences in face length and that increase the risk of suffering from BOAS. To do this, computer tomography (CT) scans of 84 different dog breeds were analysed to measure the changes in the shape of their skulls. Comparing the changes in these measurements to changes in the dogs' DNA revealed that the flat-faced dogs have an error in the *SMOC2* gene. It is thought that this error reduces the amount of the *SMOC2* protein in flat-faced breeds. This protein is normally found in the parts of the embryo that will later become the face and when it is removed from mice, it causes the mice to have abnormal head shapes similar to the flat-faced dogs. These studies strongly suggest that *SMOC2* has an important role in determining head shapes across different animals.

The Norwich Terrier is not a flat-faced dog breed, however it is known to suffer from breathing problems very similar to dogs diagnosed with BOAS. The symptoms of affected Norwich Terriers were assessed using an endoscopic camera which can examine the airways of the dogs. Changes in the airways visualised by the camera were compared to changes in the dogs' DNA, which identified a mutation in a gene called *ADAMTS3*. Exactly how this mutation causes breathing problems in the Norwich Terrier is unclear, but it may cause swelling in the airway which makes

breathing difficult. Eighty-nine different dog breeds were checked to determine if they also have the same error in the *ADAMTS3* gene. The error was only observed in the Norwich Terrier and flat-faced breeds such as the Bulldog and French bulldog. This suggests that *ADAMTS3* might be contributing to breathing problems across several breeds.

This study has found errors in DNA believed to cause the flat-faced look and breathing problems in the dog. This provides an opportunity to offer DNA tests to reduce the likelihood of these breeds suffering from BOAS and to increase the wellbeing of dogs.

Abstract

The brachycephalic head conformation is a morphological trait under human selection across many popular breeds such as the Pug. The conformation is characterised by the concurrent rostrocaudal shortening and mediolateral widening of the skull. Overwhelming evidence suggests this skull conformation predisposes breeds to a respiratory morbidity, Brachycephalic Obstructive Airway Syndrome (BOAS) which restricts airflow. This is a major welfare concern.

To date, very little is known regarding the genetic factors underpinning canine craniofacial variation or how such factors may influence BOAS aetiology. This project utilised computer tomography scans of 560 dogs to generate high resolution three-dimensional reconstructions of the skull. Geometric morphometric analysis permitted the separation of the confounding influence of size on shape (allometry) at an individual-level – a study design never previously utilised in dogs. Genome-wide association studies (GWAS) using viscerocranial shape changes identified a QTL on chromosome 1 associated with canine brachycephaly. Haplotype mapping refined the critical interval to encompass the SPARC-related modular calcium-binding protein 2 (*SMOC2*) gene. Leveraging haplotype mapping and whole-genome sequencing, a long interspersed nuclear element (LINE-1) was discovered within intron eight of *SMOC2*. Transcriptomic analysis revealed the presence of alternative *SMOC2* transcripts containing premature stop codons and a downregulation of the gene among brachycephalic carriers of the LINE-1. Models of phenotypic effect predicted that this structural variant explains up to 36% of the total craniofacial variation in the dog.

Despite many brachycephalic dogs being fixed for the *SMOC2* LINE-1 mutation, heterogeneity in BOAS presentation suggests additional genetic factors contribute to disease risk and presentation. To address this, a respiratory distress syndrome with remarkable similarities to BOAS was assessed in the Norwich Terrier. Endoscopic examinations of 233 Norwich Terriers graded and characterised the respiratory syndrome for the use in a GWAS. A single QTL on chromosome 13 was associated

with changes in the structure of the laryngeal saccules and cartilage of affected dogs. The underlying disease-associated haplotype encompassed the disintegrin and metalloproteinase with thrombospondin motifs 3 (*ADAMTS3*) gene which was observed to harbour a missense mutation in severely affected dogs. Screening for the variant across 89 diverse dog breeds revealed it was enriched in brachycephalic breeds, suggesting that the brachycephalic conformation alone does not completely explain the presentation of BOAS in these breeds. Together, these results provide the opportunity to improve animal welfare by offering genetic screening tests and further our understanding of the condition which is driven by both skeletal and non-skeletal factors.

Table of Contents

Table of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Abbreviations

2D	two-dimensional
3D	three-dimensional
BOAS	brachycephalic obstructive airway syndrome
bp	base pair
cDNA	complementary DNA
CFA	canine chromosome
CT	computer tomography
DICOM	Digital Imaging and Communications in Medicine
DNA	deoxyribose nucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
g	grams
G	g-force units
gDNA	genomic DNA
GWAS	genome-wide association study
INDEL	insertion/deletion
Kb	kilo-base pair
L	litre
LD	linkage disequilibrium
LINE-1	long interspersed nuclear element type 1
M	molar
Mb	megabase
mins	minutes
mL	millilitre
µL	microliter
mM	milli-molar
µM	micro-molar
mRNA	messenger RNA
MSC	mesenchymal stem cell

NCC	neural crest cell
ng	nano-grams
nM	nano-molar
OAS	obstructive airway syndrome
PC	principal component
PCA	principal component analysis
PCR	polymerase chain reaction
QTL	quantitative trait locus
RNA	ribonucleic acid
rpm	rotations per minute
RT-qPCR	real-time quantitative PCR
s	seconds
SDS	sodium dodecyl sulphate
SNP	single nucleotide polymorphism
Tris	tris(hydroxymethyl)aminomethane
UAS	upper airway syndrome

Foreword

The experiments within this thesis contributed to the production of two manuscripts. Firstly, the characterisation of canine skull morphology, an association of a structural variant in *SMOC2* with brachycephaly and its biological characterisation were published in *Current Biology* (published: 5th June 2017). For the purposes of this thesis, the results in this manuscript are distributed across Chapters 3, 4 and 5 to permit a significant extension and inclusion of additional experiments and discussions that could not be included in the manuscript. The content of these Chapters has been explicitly written for this thesis and are distinct from the published manuscript. All figures in the manuscript were generated exclusively by Thomas Marchant and have been modified for the inclusion in the thesis. A full copy of the Current Biology publication can be found at https://www.cell.com/current-biology/pdfExtended/S0960-9822(17)30502-X.

A second manuscript describing the causes of upper airway syndrome in the Norwich Terrier was submitted to the journal, *PLoS Genetics* (submitted: 23rd July 2018). The experiments and manuscript were executed and prepared largely by Thomas Marchant with analytical and written contributions from colleagues which are highlighted in Chapter 6. The *PLoS Genetics* manuscript is included in its entirety in Chapter 6 with extended supporting information and discussions written explicitly for this thesis.

Over the course of this thesis, conference proceedings for the Companion Animal Genetic Health (CAGH) conference 2018 were co-written and published (29th August 2018) in the journal, *Canine Genetics and Epidemiology*. Also, a lay summary of the advances in canine genetics was written for clinicians in the journal, *Veterinary Practice Today* and geometric morphometric consultation was provided for a morphological analysis of the domestic cat skull.

Publications

Marchant T.W. (2016) Harnessing Advances in Genetics to Improve Canine Health, *Veterinary Practice Today*, 4 (2), pg. 16-18.

Marchant T.W. *et al*., (2017) Canine Brachycephaly is Associated with a Retrotransposon-mediated Missplicing of *SMOC2*, *Current Biology*, 27 (11), pg. 1573-1584.

Rickets S.L. & Marchant T.W. (2018) Meeting Report from the Companion Animal Genetic Health Conference 2018 (CAGH 2018): A Healthy Companionship: The Genetics of Health in Dog, *Canine Genetics and Epidemiology*, 5 (Suppl 1).

Gordon C.R., Marchant T.W. *et al*., (2018) Morphological Variation of the Caudal Fossa of Domestic Cat Skulls Assessed with CT and Geometric Morphometrics Analysis, *Journal of Feline Medicine and Surgery*, 20 (8), pg. 752-758.

Marchant T.W. *et al*., (*In Press*) An *ADAMTS3* Missense Variant is Associated with Norwich Terrier Upper Airway Syndrome, *PLoS Genetics*.

Acknowledgements

This thesis has without doubt been one of the toughest endeavours that I have undertaken. The continued support from many colleagues, friends and family has been pivotal to its completion, and for that I am deeply grateful.

Firstly, I would like to thank my supervisors. To Jeff Schoenebeck, I am indebted to you for allowing me to join your lab and to take on this project. You have always had an open door and carved out time to meet, provided encouragement and offered unparalleled guidance, whether it's tackling a coding problem or providing recommendations for the next whisky to sample. To Dylan Clements, I am ever grateful for your fresh perspective, insightful discussions and support. Your constructive feedback and career advice has been hugely important to me.

Staying at Roslin, I would like to thank fellow Schoenebeck lab members and desk neighbours, particularly Ed, Ronan, Jenni and Anabel for your support which most often came in the form of comedic relief. Thanks to every member of the 'dog-chick' group, which at face-value was an unusual collection of disciplines but provided endless technical lab guidance and discerning discussions. Lastly, thanks to fellow students, particularly Omar and Amanda for the 'study breaks' which so often turned into overly competitive and highly animated pool competitions.

Outside of Roslin, so many flatmates, course mates and often teammates saw the less glamourous side of a PhD. There are too many people to thank in person but your continued help in separating a work from social life was decisive in finishing this project. A special thanks must also go to Neil, your friendship and principles have been, and continue to be an inspiration. Undoubtedly, I would not be in this position without them.

Finally, to my family, words can't begin to convey my gratitude for your endless patience, love and support – simply, thank you.

Chapter 1: Introduction

1.1 The Brachycephalic Skull

The mammalian skull is a profoundly complex unit which houses the visual, auditory, olfactory and gustatory sensory systems and serves to protect the brain. It also facilitates ingestion and respiration and plays a prominent role in communication and recognition. The skulls of the human, dog and mouse demonstrate that whilst the gross morphologies between species vary due to their specific evolutionary needs e.g. larger braincase in humans and more sophisticated foreface/teeth in dogs, many of their constituent bones are conserved (Figure 1.1). Each of these constituent bones (excluding the mandible) fuse together at fibrous sites called cranial sutures to form the mature cranium (Beederman *et al.* 2014). This is subdivided into the neurocranium which includes bones encompassing the brain, and the viscerocranium which includes bones of the fore-face. The comparative skull morphology across species is reflective of the shared evolutionary history of the human, dog and mouse and the high sequence synteny between them (Murphy *et al.* 2001; Arnason *et al.* 2002; Kirkness *et al.* 2003; Lindblad-Toh *et al.* 2005).

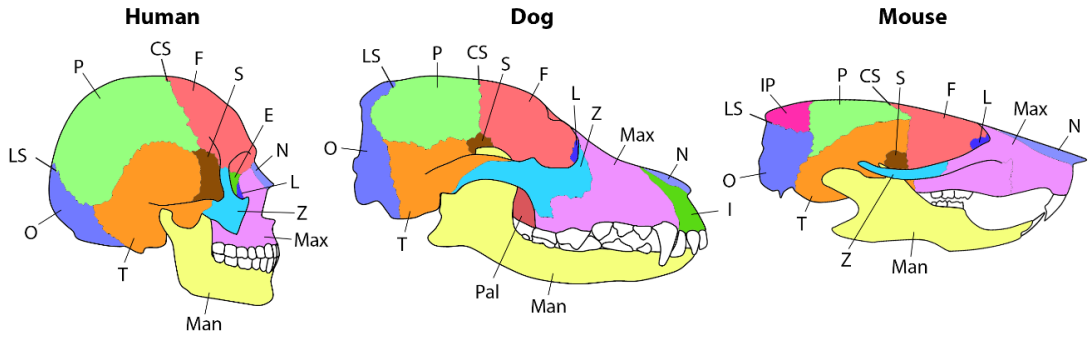


Figure 1.1 - Structure of the Mammalian Skull. Lateral view of the human, dog and mouse skulls. Bone labels: temporal (T), occipital (O), interparietal (IP), parietal (P), frontal (F), sphenoid (S), ethmoid (E), nasal (N), lacrimal (L), zygomatic (Z), incisive (I), maxilla (Max), mandible (Man) and palatine (P). The right lambdoid suture (LS) and coronal suture is (CS) is labelled with the metopic and sagittal sutures positioned along the dorsal midline of the skull between the paired parietal and frontal bones, respectively.

Morphological deviation of the skull away from the 'normal' structure for each species have been shown to induce a wide range of detrimental health problems (Wang *et al.* 2005; Bannink *et al.* 2010; Fasanella *et al.* 2010). A common collection of disorders in humans characterised by abnormal skull development are craniosynostoses.

1.1.1 Craniosynostosis

Craniosynostoses are one of the most frequent forms of craniofacial abnormalities in new-born children, predicted to affect between 2,100 and 2,500 births (Lajeunie *et al.* 1995; Persing 2008). The condition is caused by the premature fusion of one or more of the cranial sutures during the maturation of the skull which prevents further growth at that site. The remaining unfused sutures allow continued bone growth which distorts the overall shape of the skull. The sagittal suture is most commonly affected (40-58% cases), followed by the coronal suture (20-29%), metopic suture (4-10%) and lambdoid suture (2-4%) (Figure 1.1) (Kimonis *et al.* 2007). The majority (85%) of craniosynostoses occur in isolation and present clinically in a

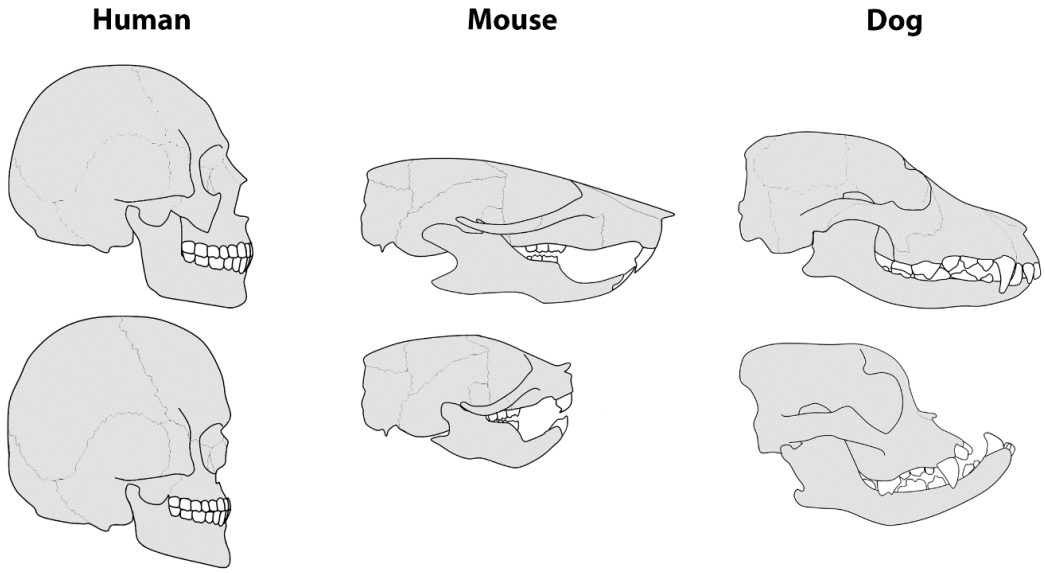


Figure 1.2 – Brachycephaly across Mammalian Species. For each of the mammalian species, a brachycephalic skull is pictured beneath that of a 'normal' skull. Brachycephalic mouse representative of the Fgfr2+/S252W mutant mouse (Wang et al. 2005). Mesocephalic and brachycephalic dogs are represented by a Husky and Boxer, respectively. Images not drawn to scale.

highly heterogenous state (Johnson and Wilkie, 2011). However, due to the finite number of possible cranial suture fusion sites, craniosynostoses can form distinct and characteristic head shapes. For example, the bilateral fusion of the coronal sutures in humans causes midface retrusion – the shortening, widening and heightening of the skull known as “brachycephaly” (Figure 1.2) (Persing *et al.* 1989). This nomenclature was used to originally describe the characteristic human head shapes but has since been adopted across species to describe similar morphological changes, whether it is the result of a disease inducing mutation, a desirable trait under selection or an intentional genetic modification (Figure 1.2) (Wang *et al.* 2005; Schlueter *et al.* 2009; Koch *et al.* 2012; Veitschegger *et al.* 2018).

A smaller proportion (15%) of craniosynostoses present as part of broader clinical syndromes which commonly have additional skeletal anomalies and most notably include abnormalities of the hands and feet (Kimonis *et al.* 2007). Crouzon, Apert and Pfeiffer syndromes are some of the most frequent forms and all have similar craniofacial characteristics including midface retrusion (brachycephaly), wide set bulging eyes, broad beaked nose, an underdeveloped jaw and are often only clinically distinguished by the extra-cranial features. Apert syndrome commonly presents with webbed or fused digits (syndactyly) whilst Pfeiffer syndrome has characteristic short digits (brachydactyly) and often has large, medially deviating thumbs and great toes (Apert 1906). The digits of Crouzon patients are largely unaffected (Anderson *et al.* 1997). Beyond these examples are over 100 further syndromic craniosynostoses but only a small subset of which have had their causal genetic variants mapped resulting in only 21% of craniosynostosis patients receiving a genetic diagnosis (Johnson and Wilkie 2011; Cox *et al.* 2014).

The structural skeletal alterations in craniosynostosis can seriously impede normal function and the quality of life of patients from an early age. Care must be taken to acutely assess restrictions to feeding, increased risk of choking and eye protection during sleep. Intracranial pressure must also be monitored to reduce the risk of developing cognitive impediments during life (Tamburrini *et al*. 2005).

Approximately 40% of patients can have intellectual developmental difficulties (Speltz *et al.* 2015) including forms of attention deficits and autism (Sarimski 1998; 2001). The brachycephalic midface retrusion in craniosynostosis also severely predisposes patients to obstructive respiratory problems such as obstructive sleep apnoea throughout life (Bannink *et al.* 2011).

1.1.2 Obstructive Sleep Apnoea

Obstructive sleep apnoea (OSA) is a chronic respiratory condition caused by the repetitive partial or complete collapse of the upper airway during sleep (Lattimore *et al.* 2003). Sufferers commonly experience fragmented sleep and oxygen desaturation and are linked with the failure to thrive, delayed development, recurrent infections, increased risk of morbidities such as cardiovascular disease and sudden death (Sin *et al.* 1999; Lattimore *et al.* 2003; Nixon and Brouillette 2005).

The midface retrusion of patients with craniosynostosis is considered to be the primary predisposing factor in OSA which stimulates the collapse of the pharynx and tonsillar hypertrophy (Bannink *et al.* 2011). The proportion of craniosynostosis patients which suffer from OSA can be as high as 70% and largely presents before 6 years of age and continues throughout life (Bannink *et al.* 2010; 2011; Driessen *et al.* 2013). Crouzon and Pfeiffer syndromes appear to have a slightly higher incidence of OSA due to their specific skull conformation compared to other forms of craniosynostosis (Moore 1993).

OSA is not just restricted to patients with craniosynostosis. Up to 5% of the general population are also diagnosed with the condition (Ward and Marcus 1996; Bixler *et al.* 1998; Sin *et al.* 1999; Bixler *et al.* 2001). Men have a slightly greater susceptibility compared to women with obesity being the greatest predisposing factor and accounts for 24% of variation in the apnoea-hypopnea index, a method for scoring the severity of OSA (Bixler *et al.* 1998; 2001; Dempsey *et al.* 2002). Interestingly, like cases of craniosynostoses, structural variation in craniofacial morphology also predisposes the general population to OSA. However, unlike craniosynostosis, the craniofacial perturbations of the general population that are

predisposed to OSA represent natural variation and are often very subtle. This natural variation has a major genetic component as shown in studies of common facial traits between twins and across families (Hunter *et al.* 1970; Nakata *et al.* 1973; King *et al.* 1993; Savoye *et al.* 1998; Johannsdottir *et al.* 2005). The heritability of these craniofacial traits is predicted to be as high as 0.8 with some traits such as cranial height being more heritable than others (King *et al.* 1993; Johannsdottir *et al.* 2005; Carson 2006).

In the general population, a single craniometric measurement, facial depth at the height of the palate can account for up to 15% of the apnoea-hypopnea index, making it the second greatest explanatory variable after obesity (Dempsey *et al.* 2002). In a separate study, face width, eye width and mandibular length can predict OSA with a sensitivity of 86% (Lee *et al.* 2009) and correlations between maxilla length are also seen in severe OSA cases (Ferguson *et al.* 1995). More specifically, Capistrano et al., (2015) showed that members of the general population with subtle brachycephaly-like craniofacial variation had a significantly increased risk of OSA compared to the mesocephalic ("normal face") and dolichocephalic ("long faced") individuals (Capistrano *et al.* 2015). Some craniofacial assessments of OSA have been less clear with regards to specific facial shapes associated with the condition. In fact, it has been suggested that the dolichocephalic head shape is also associated with cases of OSA and may in fact present an equal risk to the brachycephalic conformation (Kikuchi *et al.* 2001; Di Francesco *et al.* 2012).

Overall, variation in craniofacial morphology significantly increases the risk OSA regardless of whether it is part of a clinically diagnosed syndrome or subtle sub-clinical natural variation in the general population (Ferguson *et al.* 1995; Dempsey *et al.* 2002; Bannink *et al.* 2011). In order to understand how obstructive respiratory conditions manifest, a better understanding of the developmental origins of the skull, its genetic regulation and sources of variation is warranted.

1.2 Developmental Origins of Skull Development

Craniofacial development is a complex three dimensional (3D) process and although the duration of gestation varies between mammalian species, for example 280 days in humans and just 65 days in dogs, many of the processes are comparable (Evans and de Lahunta 2013). Integral to the development of the craniofacial skeleton is the production and migration of mesenchymal stem cells (MSC). MSCs are derived from two tissue origins; the neural crest cell (NCC) and paraxial mesoderm which are both generated during the formation and closure of the neural tube along the dorsal axis of the embryo at approximately stage 10 (early week 4) in humans (O'Rahilly and Müller 2007) (Figure 1.3A). The NCC and paraxial mesoderm populations stream into the developing frontonasal prominence and pharyngeal arches (swellings) on the ventral surface of the embryo (Figure 1.3B). In humans, there are four pharyngeal arches which begin to emerge during week four of development and contribute to different aspects of the face. Pharyngeal arch (PA) 1 has two nodules – the maxillary and mandibular prominences which give rise to the upper and lower jaw, respectively (Figure 1.3C). PA2 and PA3 most notably form the hyoid bone which acts as an anchor site for muscles of the neck, tongue and larynx. PA4 forms the larynx and the cricoid cartilage (Moore and Persuad 2003; Carlson 2009; Schoenwolf *et al.* 2015). Comparatively, in its shorter gestation, the maxillary and mandibular processes are first distinguishable at Day 23 in the dog (Pretzer 2008).

In humans, at approximately Day 35 of development the frontonasal prominence folds to form a pair of nasal pits which creates the lateral and medial nasal processes (Figure 1.3C). The medial nasal processes will merge at the midline to generate the nose, philtrum (groove of the upper lip) and primary palate whilst the lateral nasal processes form the sides of the nose. The paired maxillary processes fold and fuse to make the cheeks and maxilla whilst the mandibular processes fuse caudally to create the lower jaw and to establish the oral cavity. The final facial plan is complete by approximately week ten of development (Helms *et al.* 2005).

The bone of the skull is derived from MSCs which are distributed amongst the mesenchyme of the frontonasal prominences and pharyngeal arches. These MSCs have two distinct origins along the dorsal axis of the embryo – the NCC and paraxial mesoderm and generally contribute to the formation of the viscerocranium and neurocranium, respectively.

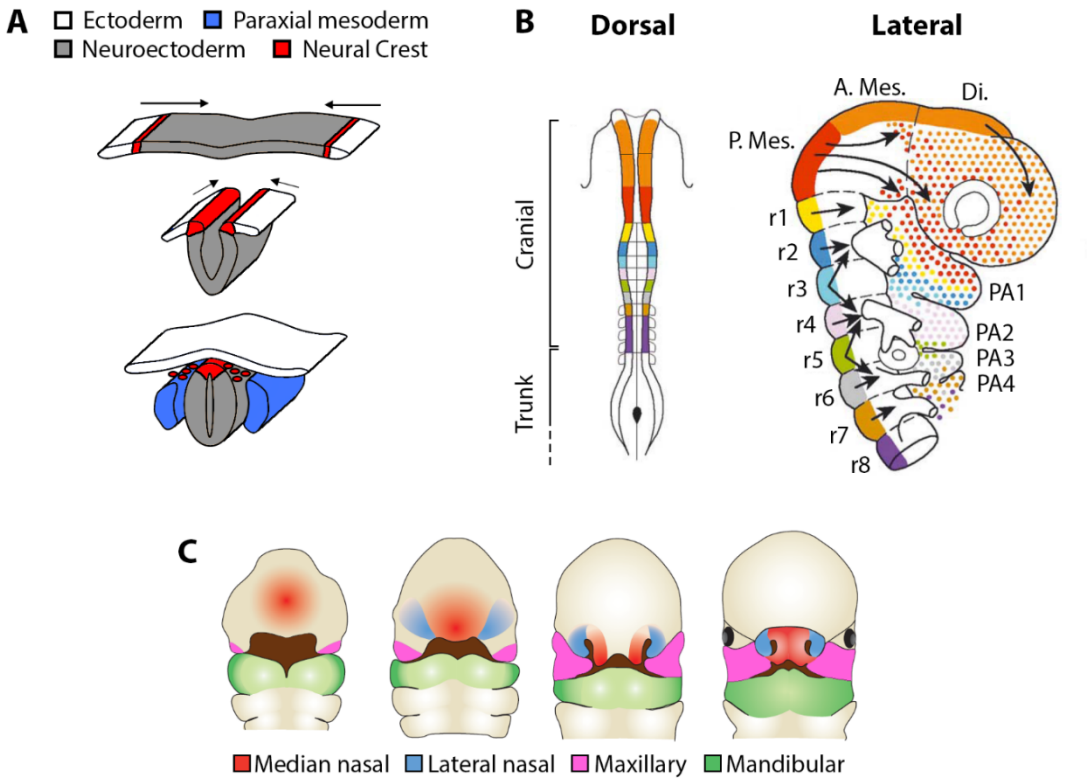


Figure 1.3 – Embryology of the Face. (A) The neuroectoderm is positioned along the dorsal midline of the embryo which invaginates to form the neural tube. Cells along the ridge of this fold form the neural crest along the length of the anterior-posterior axis. Paraxial mesoderm forms on either side of the neural tube. Both the NCC and paraxial mesoderm are sources of MSC. (B) Dorsal and lateral views of the embryo highlighting the subdivisions of the cranial neural crest by colour into the rhombomeres (r), diencephalon (Di.) and anterior/posterior mesencephalon (A./P. Mes.). NCC migrate in streams (black arrows) to populate the facial prominences and pharyngeal arches (PA). Coloured dots indicate the approximate origin of MSC at that location as predicted by fate map experiments in the chick and mouse. (C) Growth of the facial prominences between weeks 4 and 5 of the human embryo. B: modified from (Le Douarin, 2004). C: modified from (Helms et al. 2005).

1.2.1 Bones of Neural Crest Origin

The NCC is a transient population of multipotent migratory cells found exclusively in vertebrates. It is localised to the border between the neuroectoderm and ectoderm during the formation and closure of the neural tube along the length of the anterior-posterior axis (Figure 1.3A). NCC migrate around the time of neural tube closure and maintain their multipotent capacity whilst migrating (Nichols 1981; Erickson and Weston 1983; Calloni *et al.* 2009). NCCs were first identified as far back as 1868 by Wilhelm His and have since been shown to be the progenitors to many starkly diverse cell types. Cranial NCC, found at the most anterior aspect of the developing embryo have the potential to differentiate into neurons, their supporting glia, melanocytes, adrenomedullary cells, osteocytes and chondrocytes (Bronner-Fraser and Fraser 1988, Calloni et al 2009). Owing to its versatility and wide cellular potential, the neural crest has often been likened to being the fourth embryonic germ cell layer (Hall 2001).

The ectoderm-derived NCC was suggested to be involved in the formation of craniofacial mesenchyme as early as 1888 which went against the then dogma that the mesoderm was the only source of bone (Katschenko 1888). Since then, extensive work across species has demonstrated that the viscerocranium is in fact almost exclusively derived from cranial NCCs. Early cell labelling techniques used tritiated thymidine, a radioactively tagged nucleoside to track the dividing NCC as they migrated to medial and lateral nasal processes, in addition to the maxillary process of the chick (Johnston 1966; Noden 1975). Fluorescent markers such as Dil similarly tracked NCC into facial tissues and identified three well characterised streams of NCC migrating from the dorsal surface into the developing head (Lumsden 1991, Le Douarin 2004). The trigeminal stream originating from the most cranial aspect of the NCC migrates to populate the frontonasal prominence and first pharyngeal arch (Figure 1.3B). The hyoid stream emerges at the level of the fourth rhombomere and populates the second pharyngeal arch whilst the postotic stream emerges from the sixth rhombomere to populate the remaining pharyngeal arches (Figure 1.3B) (Lumsden *et al.* 1991; Trainor *et al.* 2002; Noden and Trainor 2005). These studies

were limited by the dilution of tagging markers over progressive cell divisions but gave sufficient time to grossly track NCC contributions.

During the early 1970's quail-chicken transplantation experiments became popular as a means to track cell populations throughout the course of embryonic early development using a quail-specific nuclear marker (Le Douarin 1973). Grafts of different cranial NCC regions were transplanted from the quail into the chicken host to determine the contribution of each level to the final facial bones (Figure 1.3B) (Le Lièvre 1978; Couly *et al.* 1993).

Equivalent studies in mammals were more challenging, largely due to the inaccessibility of embryos for grafting. Instead, conditional transgenic mouse lines were used to elucidate the origins of facial bones. One example system was created by crossing a mouse line expressing Cre-recombinase under the Wnt1 promoter with a second line carrying the ROSA26 conditional reporter (R26R) (Chai *et al.* 2000; Jiang *et al.* 2002). The conditional marker is only activated on the presence of Cre-recombinase which is restricted to the NCC lineage by the NCC-specific Wnt1

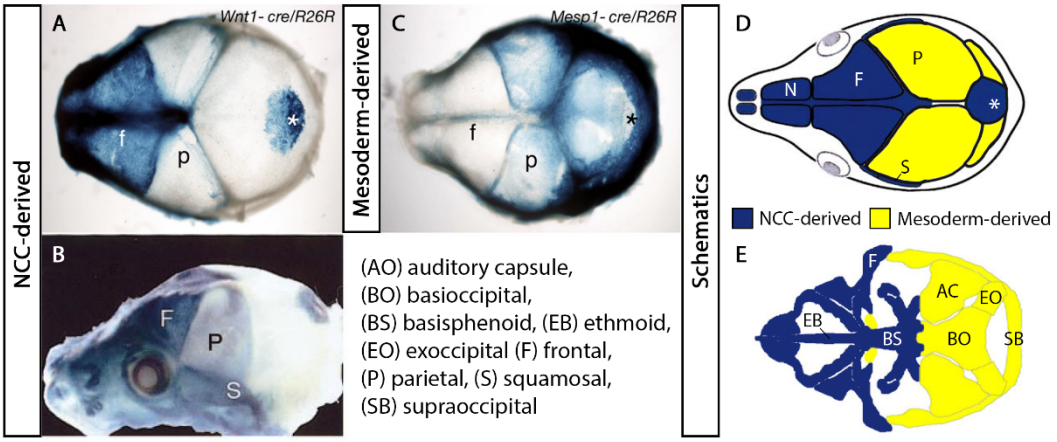


Figure 1.4 – Cellular Origins of the Bones of the Murine Skull. (A) Dorsoventral and (B) lateral views of the Wnt1-cre/R26R reporter line with the (C) Mesp1-cre/R26R reporter line showing the viscerocranium and neurocranium are derived from neural crest cell and paraxial mesoderm populations, respectively. Schematics of the (D) dorsoventral and (E) ventrodorsal views of the skull summarising the origins of cranial bones. A&C modified from (Yoshida et al. 2008). B&D modified from (Jiang et al. 2002). E modified from (McBratney-Owen et al. 2008).

promoter. All derivatives of the NCC line expressed ß-galactosidase which is identified with X-gal staining. This system confirmed that like the chicken experiments over 25 years earlier, the dorsally positioned NCC stream into populate the frontonasal prominence, the first pharyngeal arch (consisting of the maxillary and mandibular processes) and second pharyngeal arch (Chai *et al.* 2000; Jiang *et al.* 2002). These papers only studied the superficial bones of the foreface and determined that bones of the viscerocranium (nasal and maxilla), the paired frontal bones and sphenoid bones were derived from NCC cells (bones in blue) (Figure 1.4A and B) (Noden and Trainor 2005). Interestingly, the medial aspect of the interparietal bone at the posterior of the skull is also NCC-derived. It is unclear whether similar observations would be seen in the dog or human since the intraparietal bone is only distinguishable in rodents – it becomes fully fused with the occipital bone in humans and dogs (Koyabu *et al.* 2012). Of note, the Wnt1-Cre mouse line has been shown to cause ectopic expression of *Wnt1* which alters the development of the mid-brain (Lewis *et al.* 2013). It is unclear whether this would have any role on the effectiveness in tracking NCC for bone development, however a corrected Wnt1-Cre line has been made available for future studies (Lewis *et al.* 2013).

1.2.2 Bones of Mesoderm Origin

A similar reporter line was used to identify mesoderm-derived bones of the skull, this time using the mesoderm-expressed Mesp1 promoter (Jiang *et al.* 2002; Yoshida *et al.* 2008). These studies elegantly used both the Wnt1 and Mesp1 reporter lines in a complementary manner to trace the origins of the cranial vault and its underlying cranial base (chondrocranium). These experiments concluded that the paired parietal bones and the lateral aspect of the interparietal bones were derived from mesoderm (Figure 1.4C). With the exception of the medial intraparietal bone, the coronal suture acts as a border between the NCC-sourced anterior of the skull and mesoderm-sourced posterior of the skull (Figure 1.4D).

The Wnt1 and Msp1 reporter lines also identified a similar, broad anterior-posterior divide in the origin of bones of the cranial base (McBratney-Owen *et al.*

2008). Anterior bones of the chondrocranium including the ethmoid, presphenoid and basisphenoid are derived from the NCC whilst the bones positioned posteriorly to the basisphenoid are mesoderm-derived and include the auditory capsule, basioccipital, exoccipital and supraoccipital bones (Figure 1.4E). Interestingly, the basisphenoid bone appears to largely be derived from NCC however there are also mesodermal contributions (Figure 1.4E). Bones of the chondrocranium are formed by the fusion of multiple smaller cartilages prior to ossification. The constituent cartilages of the basisphenoid bone have duel origins.

1.2.3 Bone Mineralisation

In the dog, ossification of bone occurs sequentially starting rostrally in the mandible, maxilla, palatine and frontal bones at Day 35 followed by the basioccipital and basisphenoid bones and finally the caudally positioned optic capsule (Evans 1958; Watson *et al.* 1986). Regardless of the time of ossification, the skull is formed of two types of bone with distinct matrix and cellular compositions (Scott and Hightower 1991; Hayashi *et al.* 2014). The first is endochondral bone which is the major bone type throughout the body. It requires an intermediate cartilage framework to become ossified through a process called endochondral ossification. This occurs in many of the bones of the cranial base, collectively known as the chondrocranium (McBratney-Owen *et al.* 2008). The second bone type is intramembranous bone which is found almost exclusively in the skull (with the exception of the clavicle in humans). Intramembranous ossification does not need a cartilage intermediate, instead, it is directly derived from mesenchymal condensates in a process called intramembranous ossification. This produces the bones of the viscerocranium and remaining bones of the neurocranium.

Common to both types of ossification is the progression of MSCs to terminally differentiated bone-forming osteoblasts. These MSCs are derived from either the migratory NCC or paraxial mesoderm which feed into the frontonasal process and pharyngeal arches. The cellular intermediates and pathways taken between MSC and osteoblast have similarities but ultimately differ according to ossification type. MSCs

have the capability to give rise to chondrocytes, osteoblasts, myoblasts and adipocytes which are under the control of master transcription factors (Rutkovskiy *et al.* 2016). SRY (Sex-Determining Region Y)-Box 9 Protein (*Sox9*) and runt-related transcription factor 2 (*Runx2*) determine the fate of MSCs by initiating endochondral or intramembranous ossification, respectively (Figure 1.5) (de Crombrugghe *et al.* 2000; Chen *et al.* 2014).

1.2.3.1 Endochondral Ossification

During endochondral ossification, the cartilaginous fate of mesenchymal cells is induced by the *Sox9* transcription factor, under bone morphogenetic protein (BMP) control (Bi *et al.* 1999; Healy *et al.* 1999; de Crombrugghe *et al.* 2000; Nishimura *et al.* 2012). This stimulates their differentiation into chondrocytes and expression of classic chondrocyte genes such collagen II and aggrecan (Chambers *et al.* 2002) (Figure 1.5). Chondrocytes cease to grow and undergo a condensation step via an N-cadherin dependent mechanism on their cell surface (Janners and Searls 1970; Oberlender and Tuan 1994). Indian hedgehog homologue (IHH) and parathyroid hormone-related protein (PTHrP) signalling stimulate chondrocytes to proliferate rapidly and secrete a variety of matrix proteins to generate the intermediate cartilage

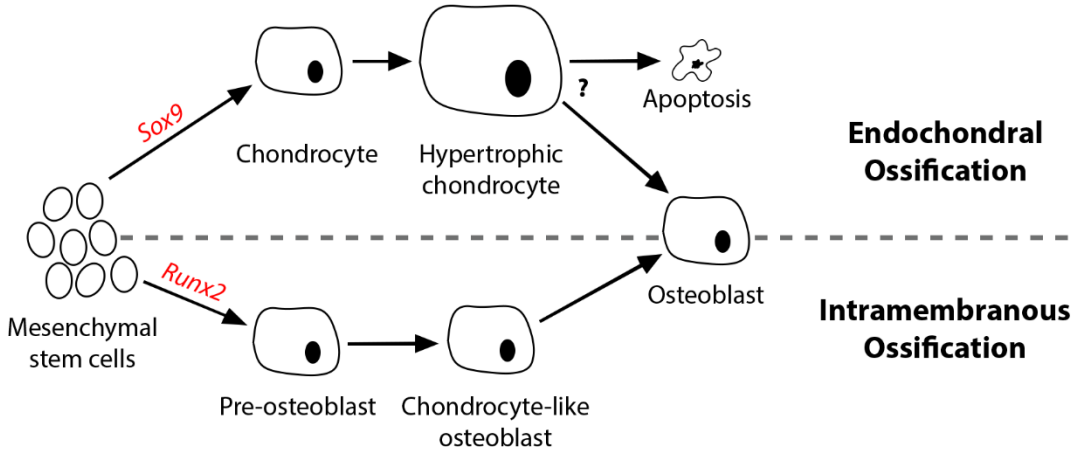


Figure 1.5 – Bone Formation of the Skull. Two methods of bone formation exist in the skull. Generally, posterior bones of the skull are produced through endochondral ossification which requires chondrocytes to build a cartilage template prior to mineralisation by osteoblasts. Meanwhile, anterior bones of the skull are produced through intramembranous ossification which directly forms bone from mesenchymal condensates.

structure (St-Jacques *et al.* 1999; DeLise *et al.* 2000; Zhao *et al.* 2002). Following the formation of the cartilage, the chondrocytes swell over 20-fold in size to form hypertrophic chondrocytes (Cooper *et al.* 2013). This terminally differentiated chondrocyte marks the onset of bone formation as the cells secrete collagen X and fibronectin which permits the calcification of the matrix (Zheng *et al.* 2003). The fate of hypertrophic chondrocytes is unclear with some reports claiming they undergo apoptosis whilst others suggest they may go on to transdifferentiate into bone-forming osteoblasts (Gibson 1998; Yang *et al.* 2014). Regardless of their fate, vascularisation of the cartilage framework accompanied by the invasion of osteoblasts and osteoclasts permits the deposition of a mineralised bone matrix and establishment of bone marrow (Erlebacher *et al.* 1995). Eventually the cartilage framework is completely replaced by bone.

1.2.3.2 Intramembranous Ossification

The intramembranous ossification process is initiated by the master transcription factor *Runx2* (Figure 1.5) (Abzhanov *et al.* 2007; Takarada *et al.* 2015). Bmp2/4 then act in a threshold-dependent manner to confirm the fate of the MSCs. Low levels permit intramembranous ossification via condensation and differentiation into pre-osteoblast cells however high Bmp levels stimulate the formation of chondrocytes (see previous) (Liu *et al.* 2005; Abzhanov *et al.* 2007; Bonilla-Claudio *et al.* 2012). For intramembranous ossification, the Bmp signalling regulates the self-renewal and proliferation of pre-osteoblast cells (Bonilla-Claudio *et al.* 2012). Signalling peptides PTHrP and IHH then regulate the differentiation of pre-osteoblast cells into chondrocyte-like osteoblasts (Lenton *et al.* 2011) – so named because they express both chondroblastic (collagen II/IX) and osteoplastic (osteopontin) markers simultaneously (Abzhanov *et al.* 2007). Finally, the chondrocyte-like osteoblasts differentiate into the mature osteoblast. The osteoblasts secrete an osteoid matrix containing osteopontin, alkaline phosphatase and collagen I which mineralises and encases the osteoblasts within (Stein *et al.* 1989; Jabalee *et al.* 2013). These trapped osteoblasts have the capability to become osteoclasts which together regulate bone formation (Franz-Odendaal *et al.* 2005).

1.3 The Genetic Determinants of Craniofacial Variation

Given that craniofacial development consists of a complex interplay of migrating, proliferating and differentiating cells from multiple origins under the tight coordination of numerous major signalling pathways, it is perhaps unsurprising that subtle perturbations in any of these steps give rise to craniofacial variation. This variation can vary from subtle changes in the position/depth of facial features to substantial alterations that present significant clinical challenges whether in human or veterinary medicine.

1.3.1 Causes of Craniosynostoses

The discoveries to date of causal variants of craniosynostoses were predominantly identified in rare or family-based studies. This included the first craniosynostoses (Boston-type) to be mapped to a pair of missense mutations in the Msh homeobox 2 (*MSX2*) gene which are predicted to disrupt neural crest cell migration and survival (Jabs *et al.* 1993; Liu *et al.* 1995; Ma *et al.* 1996; Winograd *et al.* 1997; Finnerty *et al.* 2009). Meanwhile, mutations within the twist-related protein 1 (*TWIST1*) were shown to cause Saethre-Chotzen syndrome by stimulating the premature maturation of bone-forming osteoblasts in cranial bones (Howard *et al.* 1997; Yousfi *et al.* 2002; Bialek *et al.* 2004; Guenou *et al.* 2005). A number of further rare craniosynostoses were shown to be caused by missense mutations in genes such as Gli family zinc finger 3 (*GLI3*, Greig syndrome), establishment of cohesion 1 homolog 2 (*ESCO2*, Robert syndrome) and Jagged 1 (*JAG1*, Alagille syndrome) (Table 1.1) (Wild *et al.* 1997; Colliton *et al.* 2001; Vega *et al.* 2005).

Beyond these, some of the more common syndromic craniosynostoses such as Crouzon, Apert and Saethre-Chotzen syndrome have a remarkably narrow aetiology. Many of the mutations identified in these conditions are restricted to the fibroblast growth factor receptor (*FGFR*) family and its related pathways. Of the diagnosed cases, mutations have been found in *FGFR2* (32%), *FGFR3* (25%), *TWIST1* (19%) and Ephrin B1 (*EFNB1*) (7%) (Johnson and Wilkie 2011). Interestingly, variants within the *FGFR2* gene have been shown to cause a number of different forms of

craniosynostoses and in some instances the exact same mutation is responsible (Reardon *et al.* 1994; Jabs *et al.* 1994; Rutland *et al.* 1995; Wilkie *et al.* 1995; Oldridge *et al.* 1999; Bochukova *et al.* 2008; Chokdeemboon *et al.* 2013) (Table 1.1). The restriction of a collection of craniosynostoses with similar craniofacial abnormalities to a narrow network of affected genes is perhaps unsurprising. However, given the variable extra-cranial phenotypes in the hands and feet of many of these syndromes, it likely suggests a more complex relationship between these variants and other genetic loci.

In order to further elucidate the underlying genetics of complex traits such as craniofacial morphology, genome-wide association studies have been a popular study design across fields in recent years (Thorleifsson *et al.* 2008; Marouli *et al.* 2017). To date, only one study has sought to use GWAS for craniosynostoses. Specifically, the study focussed on a non-syndromic form of craniosynostosis affecting just the sagittal suture and found significant associations in regions adjacent to the *BMP2* and Bardet-Biedl syndrome protein 9 (*BBS9*) genes (Justice *et al.* 2012).

Gene	Condition	Reference
FGFR1	Pfeiffer syndrome	(Muenke *et al.* 1994; Chokdeemboon *et al.* 2013)
FGFR2	Apert syndrome Crouzon syndrome Jackson-Weiss syndrome Pfeiffer syndrome	(Reardon *et al.* 1994; Jabs *et al.* 1994; Rutland *et al.* 1995; Wilkie *et al.* 1995; Oldridge *et al.* 1999; Bochukova *et al.* 2008; Chokdeemboon *et al.* 2013)
FGFR3	Crouzon syndrome	(Meyers *et al.* 1995; Muenke *et al.* 1997)
ALX3	Frontonasal syndrome	(Twigg *et al.* 2009)
EFNB1	Frontonasal syndrome	(Wieland *et al.* 2004; Twigg *et al.* 2004)
ESCO2	Roberts syndrome	(Vega *et al.* 2005)
GLI3	Greig syndrome	(Wild *et al.* 1997)
JAG1	Alagille syndrome	(Warthen *et al.* 2006)
MSX2	Boston-type	(Jabs *et al.* 1993; Ma *et al.* 1996; Florisson *et al.* 2013)
TWIST1	Saethre-Chotzen syndrome	(Ghouzzi *et al.* 1997; Howard *et al.* 1997)

Table 1.1 – Mutations Causing Craniosynostoses. Summary of the genes harbouring mutations known to cause some of the most common forms of craniosynostosis.

Later, a variant in a regulatory element of *BMP2* was shown to be the cause of the preceding association study (Justice *et al.* 2017).

These studies begin to broaden the pool of known candidate genes and causal variants for craniosynostoses however it remains that just 21% of patients with craniosynostoses receive a genetic diagnosis (Johnson and Wilkie 2011). Due to the rarity of clinical cases, the recruitment of adequate populations to power such GWAS is extremely challenging as reflected by only the one GWAS-based craniosynostosis study to date. Therefore, an alternative approach to determining the genetic determinants of craniofacial variation has been to focus on the natural, sub-clinical variation that exists in the human general population.

1.3.2 Associations with Normal Variation

The stark craniofacial alterations in craniosynostoses permits for the use of a case vs control phenotype for GWAS. However, the subtle morphological variation amongst the general population requires a more accurate means to capture and decipher the craniofacial form as a phenotype. Recent developments and increased availability of several precision phenotyping techniques coupled with cheaper SNP microarray data have permitted large-scale assessments with as many as 5,300 participants. Analyses adopting GWAS have focused on methods to capture the 3-dimensional (3D) shape of the face, using superficial tissues as landmarks e.g. edges of the nose, mouth, eyes etc. A commonly adopted approach has been to overlay multiple 2D images taken at multiple angles to generate a dense 3D surface reconstruction of the face (known as stereophotogrammetry) (Hammond *et al.* 2004). Subsequently, facial features are landmarked using geometric morphometrics to give 3D coordinate-based phenotypes. This technology has advanced significantly from it first use and now permits fully automated landmarking of large reconstruction datasets, which negates the previous need of manual landmarking samples which was time consuming and prone to user error (Guo *et al.* 2013).

To date, five large-scale GWAS exploring normal craniofacial variation in humans have identified associations surpassing genome-wide significance.

Paternoster et al., 2012 first adopted this stereophotogrammetric approach to study a UK population of adolescents (Paternoster *et al.* 2012). Their association studies using distances between 3D landmarks identified a genome-wide significant marker within the Paired Box Protein 3 (*PAX3*) gene to be associated with variation in the nasion position (Table 1.2). *PAX3* makes for an excellent candidate for modifying craniofacial features. It is a transcription factor expressed and enriched in cranial neural crest cells and neural crest-derived craniofacial mesenchyme (Goulding *et al.* 1991; Hagiwara *et al.* 2014). *PAX3* regulates the balance between proliferation and differentiation of the neural crest lineage (Monsoro-Burq 2015). Over 80 mutations within this gene have been linked with Waardenburg syndrome, which amongst other defects, commonly presents with the widening of the nasal bridge (location of the nasion) (Pingault *et al.* 2010). Interestingly, a persistent overexpression of *PAX3* causes cleft palates in mice (Wu *et al.* 2008).

Author	Gene	Phenotype
Paternoster *et al*. 2012	*PAX3*	Nasion position
Liu *et al*. 2012	*PAX3*	Nasion-eye distance
	PRDM16	Nose width & height
	TP63	Distance between eyes
	COL17A1	Nasion-eye distance
Adhikari *et al*. 2016	*DCHS2*	Nose inclination
	RUNX2	Nose width (bridge)
	GLI3	Nose width (wing)
	PAX3	Nasion position
	EDAR	Chin protrusion
Cole *et al*. 2016	*SCHIP1*	Face size
	PDE8A	Face size
Lee *et al*. 2017	*FREM1*	Lip height
	PARK2	Midface height

Table 1.2 – Positional Gene Candidates of Normal Human Variation. Summary of positional gene candidates identified across genome-wide association analyses of normal human craniofacial variation. Gene candidates are only listed if significant markers are located within a gene.

PAX3 was then subsequently identified in a further two independent studies which also used stereophotogrammetry or an magnetic resonance image (MRI) scanner to capture the surface structure of participants (Liu *et al.* 2012; Adhikari *et al.* 2016). Together, these studies also found associations with variations in craniofacial features with *RUNX2*, *GLI3*, PR domain containing 16 (*PRDM16*), transformation-related protein 63 (*TP63*), collagen type XVII alpha 1 chain (*COL17A1*), dachsous cadherin-related 2 (*DCHS2*) and ectodysplasin A receptor (*EDAR*) (Table 1.2). Two further stereophotogrammetry-based studies found associations with the schwannomin interacting protein 1 (*SCHIP1*), phosphodiesterase 8A (*PDE8A*), FRAS1-related extracellular matrix 1 (*FREM1*) and parkin 2 (*PARK2*) genes (Cole *et al.* 2016; Lee *et al.* 2017) (Table 1.2). Notably, many of these variants are predicted to have low effect sizes with limited heritability. In an attempt to find large effect variants, a recent study by Crouch *et al.* first identified highly heritable regions of the face using a cohort of twins prior to using extreme variations of these regions in the People of the British Isles study to execute a GWAS (Crouch *et al.* 2018). This study identified variants in the protocadherin-15 (*PCDH15*), transmembrane protein 163 (TMEM163), and membrane-bound transcription factor site-1 protease (*MBTPS1*) genes.

Collectively, many of these aforementioned positional genes are compelling candidates for regulating craniofacial variation. *RUNX2* and *GLI3* are the most striking due to their known known roles in regulating bone mineralisation (see Section 1.2.3) and causing a form of craniosynostosis (Greig Syndrome), respectively (Wild *et al.* 1997; Abzhanov *et al.* 2007; Takarada *et al.* 2015) whilst knockout mice for *Frem1* and *Schip1* both have abnormal foreface shapes (Schmahl *et al.* 2006; Alazami *et al.* 2009; Slavotinek *et al.* 2011). From this, it is clear that GWAS is a powerful method for detecting genetic signatures of facial features albeit it, often finding variants of low effect size. Due to the high number of tests and potential underlying population structure, spurious associations can mistakenly be found although this can be accounted for by careful cohort selection as demonstrated by the reproducible discovery of candidate genes such as *PAX3*. Importantly, causal genetic variants at these loci were not identified and fundamentally it remains to be seen whether these

associations are representative of changes in the superficial soft tissue of the face or its underlying skeleton.

1.3.3 Associations with Obstructive Sleep Apnoea

In contrast to the exploration of craniofacial variation using GWAS, investigations into the underlying genetics of OSA have predominantly used association studies with candidate genes. A major advance of using this study design is to significantly increase the power of detecting associations with small size effects (Risch and Merikangas 1996). To this end, measurements of OSA such as the apnoea-hypopnoea index have been associated with gene markers within c-reactive protein (*CRP*), glial cell-derived neurotrophic factor (*GDNF*) and 5-hydroxytryptamine receptor 2A (*HTR2A*) amongst others (Larkin *et al.* 2010; Patel *et al.* 2012; Grilo *et al.* 2013; Xu *et al.* 2014; Kripke *et al.* 2014; Baik *et al.* 2015; Cade *et al.* 2016). However, in line with the common criticisms of candidate gene studies, these analyses have failed to give reproducible findings across studies (Tabor *et al.* 2002). Also, such studies fundamentally rely on prior knowledge and an understanding of pathways and genes that are potentially involved in disease phenotypes (Tabor *et al.* 2002). These studies have relied on databases of gene candidates associated with the circadian rhythm, obesity, inflammation, ventilation control, heart defects and blood disorders which are potential target networks but ultimately introduce an element of bias and negate the discovery of novel genes and networks (Larkin *et al.* 2010; Patel *et al.* 2012; Grilo *et al.* 2013; Xu *et al.* 2014; Kripke *et al.* 2014; Baik *et al.* 2015).

From the collective studies on craniosynostosis, normal craniofacial variation and OSA, it remains that just 21% of craniosynostoses receive a genetic diagnosis and no known variants cause OSA. The recent adoption of GWAS-focussed study designs in the aforementioned analyses have demonstrated the effectiveness of association studies in detecting genetic signatures of facial variation in humans. Despite this, studies in humans are limited by the rarity of extreme craniofacial phenotypes, such as those in craniosynostoses and the subtly of 'normal' non-clinical variation which requires vast population cohorts and highly accurate approaches to phenotyping. For

this reason, the domestic dog was chosen to investigate the underlying genetics of craniofacial diversity and presentation of severe obstructive respiratory conditions to overcome the challenges in the human field. The dog has highly diverse craniofacial features with comparable morphologies to the brachycephalic and dolichocephalic skull conformations seen in humans. Once more, these extreme morphologies are highly prevalent in the UK dog population and many of which are susceptible to obstructive respiratory conditions with parallels to OSA. Beyond this, the dog has a simplified genetic architecture which makes it a highly advantageous system to work with.

1.4 The Dog as a Genetic Research Model

Several characteristics unique to the dog make the genetic endeavours of morphological and disease trait mapping in the species much more appealing than others. A notable advantage of the domestic dog is its popularity. In 2018, over 1 in 4 (26%) UK households have a pet dog and over £4.2 billion was spent on veterinary healthcare in 2017, which was an 18.7% increase on the previous year (Office for National Statistics 2018; PFMA 2018). As a result, there is an unprecedented level of access to demographic information, diagnostic test results and biological samples over the course of a dog's life (Patterson 2000). Also, throughout the course of their lifetime, dogs age comparatively to humans, albeit over an approximately 7-8 times shorter period whilst being exposed to the same environmental factors. This enables us to draw much better parallels between human and canine disease presentation, progression and response to treatments unlike many of the artificially introduced disease states in classic model organisms such as the mouse (Khanna *et al.* 2006; Breen and Modiano 2008). Over 360 of the approximately 450 recognised diseases in the dog have equivalents in humans which include complex traits such as diabetes and cancers (Shearin and Ostrander 2010). Many of these diseases were shown to be caused by gene homologues across the two species making the dog a powerful comparative disease model (Ostrander *et al.* 2000).

Although genetic investigations into the domestic dog would appear less common-place compared to other species such as humans, mice and zebrafish, there is a growing availability of genomic tools for the dog. This started with the availability of the first 1.5X assembled Poodle reference sequence in 2003 followed by a 7.5X Boxer sequence in 2005 which are available for public use on the UCSC genome browser (Kirkness et al 2003; Lindblad-Toh et al 2005). The most up to date CanFam3.1 assembly of the dog genome has also been annotated (Pruitt et al. 2007; Pruitt et al. 2014). More breed-specific assemblies are currently being generated including that of a Great Dane and a Labrador which will allow for better short-read sequence alignment and study-specific analyses (personal communications). To aid in the discovery of variants, commercially available 170,000 SNP microarray chips have been made produced. Free to access repositories containing both SNP microarray data and whole genome sequences such as the Dog Biomedical Variant Database Consortium (DBVDC) and Dog Genome SNP Database (DoGSD) have been established to support variant analysis in the community (Bai et al. 2015). Many of the genomic software packages designed to analyse such databases were first set up for human studies but can easily be transferred to canine experiments. For example, association studies can be carried out in the PLINK toolset using dog-specific modifiers and arguments whilst packages such as GEMMA can be used to invoke kinship matrices to account for relatedness and population structure in dogs (Purcell *et al.* 2007; Zhou and Stephens 2012).

For traits such as morphology, the dog represents the most morphologically diverse land mammal with extremes of phenotypes e.g. there can be over a 40-fold difference in body weight between the toy Chihuahua and giant Great Dane (Stockard 1941). Such broad phenotypic scales coupled with their advantageous genomic structure ideally positions the dog to tease apart phenotype-genotype relationships. Both of these characteristics are the result of its unique evolutionary history of the dog.

1.4.1 Domestication's Impact on the Canine Genome

The modern dog, *Canis familiaris* was domesticated from the grey wolf, *Canis lupus* estimated to be within the last 11,000 to 32,000 years (Benecke 1987; Tchernov and Valla 1997; Boyko *et al.* 2009; Frischknecht *et al.* 2013; Freedman *et al.* 2014; Frantz *et al.* 2016). The domestication of the dog is thought to be a multi-event process in which small populations of wolves at various sites (most likely in South-East Asia and European areas such as the Levant) were translocated and bred together over time. The limited pool of founder wolves and subsequent bottlenecks during the establishment of the modern dog breeds ~200 years ago caused significant founder effects within the species (Wayne and Ostrander 2007; Freedman *et al.* 2014; Frantz *et al.* 2016). Later, the repeated use of popular sires and events such as the World Wars saw the breeding populations further reduced. For example, the breeding populations of the English Mastiff was reduced to just 14 and Manchester Terriers to as few as 11 (Larson *et al.* 2012). Now, 80% of chromosomes across breeds carry between just two and four haplotypes and linkage disequilibrium (LD) can extend up to 3 Mb in breeds such as the Bernese Mountain Dog (Sutter *et al.* 2004). As a result, studies found that LD within the domestic dog is approximately 100 times greater than that in humans (Ostrander and Kruglyak 2000; Sutter *et al.* 2004). Genomic alterations such as this are advantageous in that fewer genetic markers are required to tag haplotypes which is a major advantage for genotyping arrays, however the expanse of LD makes the subsequent fine mapping more challenging.

At a variant-level, there has been a marked increase in the number of nonsynonymous deleterious variants in the dog which are attributed to the relaxation of natural selection pressures (Björnerfeldt *et al.* 2006; Cruz *et al.* 2008). Further modifications include the accumulation and increased heterogeneity of structural variants within the canine genome compared to the wolf and human. Elements such as long- and short-interspersed nuclear elements (LINE/SINE) are heterozygous at over 10,000 sites in the Boxer compared to the Poodle (Kirkness *et al.* 2003; Lindblad-Toh *et al.* 2005; Wang and Kirkness 2005). Such genetic variation coupled with the extensive fluctuation of repeat element length and hypermethylation near key

coding regions of the genome are significantly enriched in the dog (Fondon and Garner 2004; Koch *et al.* 2016). The absence of these genomic signatures in the wolf and their rapid accumulation and heterogeneity in the dog suggest they could disrupt gene transcription and drive the extreme phenotypic diversity in dogs. Using this phenotypic diversity as a selection trait, successive and repetitive breeding within the confines of the breed barriers have enriched alleles harbouring these variations almost to the extent of fixation in some breeds whilst being completely absent from others. The breed-specific enrichment of desirable traits and their alleles can also inadvertently allow the co-segregation of disease states within breeds in what is known as a so-called "hitch-hiker effect". As a result, in many cases, shared morphological traits and disease states across individuals are thought to be caused by identity by descent (IBD) genetics in which the same causal variant is shared between individuals by a common ancestor. Helpfully, when resolving derived haplotypes shared by IBD genetics, recombination events can occur around the associated variant to refine the haplotype. Providing the study design is permitting, further improvements to fine mapping might be achieved through the inclusion of mixed-breed dogs

Collectively, these effects have resulted in stark phenotypic differences between breeds whilst preserving the same basic genomic background – an ideal combination to elucidate the underpinnings of morphological variation and disease.

1.4.2 Canine Craniofacial Variation

Craniofacial variation in humans is particularly modest with most variation going unnoticed without accurate methods of phenotyping. Furthermore, the instance of severe craniofacial variation is relatively rare such as the craniosynostoses which occur approximately once in 2,500 new-born children (Persing 2008). This presents a considerable challenge to investigating the underlying genetic causes of human craniofacial variation which often require very large population cohorts to do so.

In contrast, the domestic dog represents the most morphologically diverse land mammal with some of the greatest variation observed in the face (Stockard 1941). Many of the dogs' facial traits mimic the changes seen in humans, albeit on a much more extreme scale. Breeds such as the Boxer and Greyhound are representatives of the extreme brachycephalic and dolichocephalic head conformation, respectively (Figure 1.6B and D). Unlike humans, the brachycephalic trait is very common due to the popularity of breeds such as the Bulldog, French bulldog, Boxer, Pug, French Mastiff, Shih Tzu and Pekingese. The "flat-faced" trait in dogs was thought to be popularised either for its functionality by increasing bite forces needed by breeds such as the Bulldog for their original role in bull-baiting (Ellis *et al.* 2009); or aesthetically due to the pronounced eyes and enlarged forehead, characteristic of juveniles (Drake 2011; Geiger *et al.* 2017). The canine skull is also rich in other head shape traits including mandible position (e.g. mandibular prognathism in the Boxer), angle of the palate (e.g. klinorynchy in Bull Terrier) and neurocranium shape (e.g. apple vs deer head shape in the Chihuahua).

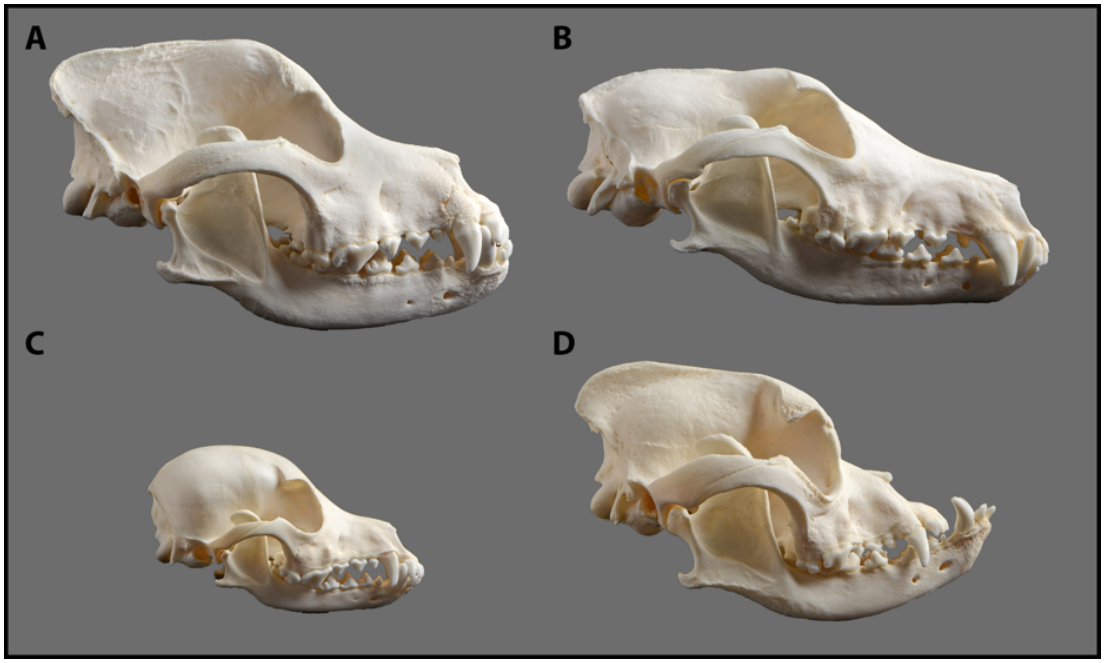


Figure 1.6 – Skull Variation. Lateral skull images of a purebred (A) Rottweiler, (B) Greyhound, (C) Miniature Dachshund and (D) Boxer representing some of the diversity across the Canis familiaris species. Photographs are to scale.

Noticeably, there is also a significant variation in the overall size of the skull between breeds such as that between the Rottweiler and Miniature Dachshund (Figure 1.6 A and C) or in more extreme cases, the Chihuahua and the Great Dane (not shown here). Such extreme morphological variation also permits the concurrent exploration of the genetic determinants of head size (and related body size). However, this extreme variation in the size of the dog can have a confounding effect on the assessment of shape changes in the skull – a phenomenon known as allometry.

1.4.2.1 The Confounding Effects of Allometry

As the size of an organism varies, many of its traits will vary proportionally on a 1:1 scale with its size – this is an isometric relationship and can be seen in the proportional increase in wing size with the increase in body size of the *Drosophila melanogaster* (Figure 1.7A) (Shingleton *et al.* 2009). Conversely, some traits do not scale proportionally relative to body size in a relationship known as allometry. An allometric relationship can describe either an increase or decrease in trait size relative to body size – positive and negative allometry, respectively (Shingleton and Frankino 2012). For example, negative allometry is observed in the brain size of humans which has a reduced sensitivity to the changes in their overall body sizes

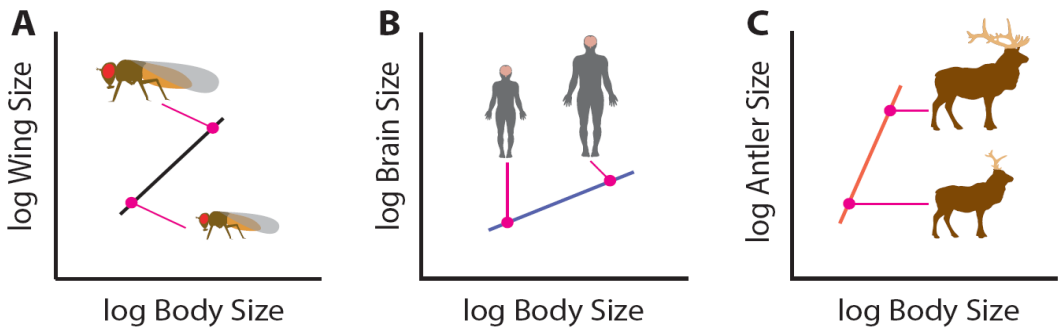


Figure 1.7 – Morphological Scaling. (A) The size of the *Drosophila* wing has a 1:1 (isometric) scale with variation in its body size. (B) The size of the human brain is less sensitive to changes in overall body size and scales disproportionately (negative allometry). (C) The antlers of the male Irish elk are hypersensitive to changes in body size and become proportionally greater in size with incremental increases in body size (positive allometry). Modified from: (Shingleton and Frankino 2012).

(Figure 1.7B) (Koh *et al.* 2005). Meanwhile the antlers of a male Irish elk are hypersensitive whereby a small increase in body size will result in a proportionally greater increase in antler size (Figure 1.7C) (Gould 1973).

To date, few studies have assessed allometric relationships of traits in the dog. One study tracked ontogenetic allometry (allometry over an individual's lifetime) in the Beagle (Helmsmüller *et al.* 2013). It found that amongst body circumference, pelvis size and foot size, the head length had a negative allometric relationship with the Beagle's body weight. Meanwhile, traits such as pelvic height, height at the withers and femur length had strong positive allometry. The observation of leg bone allometry was confirmed across 63 dog breeds in a separate study (Casinos *et al.* 1986). Although less appropriate here, two subsequent studies have explicitly focussed on the effects of allometry on sexual dimorphism across breeds albeit with conflicting conclusions. One study concluded that males of larger breeds were proportionally greater in size relative to a breed-matched female compared males of smaller breeds – a biological observation recorded across the animal kingdom known as Rensch's rule (Frynta *et al.* 2012). The second study observed no such proportional changes however it analysed far fewer breeds with smaller population numbers compared to the former (Sutter *et al.* 2008).

Nonetheless, the presence of allometry within the skull of the dog has been made evident by previous studies analysing the changes in the St Bernard overtime and modularity of the canine skull (Drake and Klingenberg 2007; 2010). Due to the extreme variation in body size of the domestic dog, the confounding influences and modifying effects an individual's body size on its variation in face shape cannot be overlooked.

1.4.3 Brachycephalic Obstructive Airway Syndrome

In conjunction with providing the opportunity to explore the genetics of the diverse craniofacial variation, the dog is also susceptible to a number of conformation-related morbidities (Hodgman 1963; Hillbertz *et al.* 2007; Asher *et al.* 2009; Karyadi *et al.* 2013). Some of the most common morbidities are linked to the

brachycephalic skull conformation and include gastrointestinal tract lesions (Poncet *et al.* 2005a; Harvey and Haar 2016), ear and eye-related morbidities (Sanchez *et al.* 2007; Packer *et al.* 2015; Harvey and Haar 2016; O'Neill *et al.* 2017), neurological conditions (Collmann *et al.* 2005), spinal/vertebral defects (Ryan *et al.* 2017; Mayousse *et al.* 2017; de Souza *et al.* 2018) and dystocia (obstructed labour) (O'Neill *et al*. 2017). Furthermore, the brachycephalic dogs have a significantly shorter lifespan and quality of life compared to their dolichocephalic counterparts (O'Neill *et al.* 2015).

In line with OSA in humans, changes in canine craniofacial morphology also predisposes the dog to a severe obstructive respiratory problem called brachycephalic obstructive airway syndrome (BOAS). BOAS is one of the most frequently reported morbidities in brachycephalic dogs such as the Pug, Bulldog, French Bulldog, Boston Terrier, Pekingese and Shih Tzu (Poncet *et al.* 2005b; Torrez and Hunt 2006; Riecks *et al.* 2007; Ginn *et al.* 2008; De Lorenzi *et al.* 2009; Fasanella *et al.* 2010).

1.4.3.1 BOAS Prevalence

As its name suggests, BOAS is a chronic respiratory condition which is almost exclusively seen in brachycephalic dog breeds. It encompasses a range of primary and secondary pathological alterations to the upper respiratory tract which increase respiratory effort, resulting in laboured breathing, intolerance to heat/exercise, cyanosis and collapse (Ginn *et al.* 2008; Bernaerts *et al.* 2008). The reported prevalence of the condition varies widely which is likely the result of the methods used in the assessment. Owner-directed questionnaire-based assessments of BOAS have predicted the condition to affect less than 5% of each brachycephalic breed (Wiles *et al.* 2017) however these investigations are markedly biased in that select UK Kennel Club registered dog owners were asked to report conditions in their dogs. Such surveys are undermined by either the unwillingness to report ill-health, or the inability of non-professionals to identify clinical symptoms. Indeed, 58% of owners with dogs clinically diagnosed with BOAS did not perceive their dogs to have any

breathing problems (Packer *et al.* 2012). Meanwhile, in veterinary practices the prevalence of BOAS is reported to be much higher and varies by breed - Bulldog (18.0 – 61.0%), French bulldog (2.7 – 67.0%), Pug (6.8 – 50.0%) and Boston terrier (1.4 – 12.9%) (Poncet *et al.* 2005b; Torrez and Hunt 2006; Riecks *et al.* 2007; Ginn *et al.* 2008; De Lorenzi *et al.* 2009; Fasanella *et al.* 2010). Again, these estimates are subject to the innate biases of using veterinary healthcare records; the majority of dogs visiting a veterinary practice are doing so because of their pet's ill-health and so the data may not truly reflect the disease prevalence across the entire population (of which healthy dogs aren't represented).

Objective assessment using whole body barometric plethysmography has found signatures of respiratory distress in up to 89% of French bulldogs, 54% of which presented with a clinically-recognised level of disease (Liu *et al.* 2015). The disparity between the perceived rates of BOAS and the underlying levels of respiratory dysfunction in this breed is profound, both from a lay and a clinical professional perspective, and likely to be transferable to other breeds. One possible explanation for this disparity in perception is the desensitisation towards the condition. Loaded terminology and phrases used when describing respiratory symptoms such as 'laboured breathing is *normal* for the breed' indicate a level of acceptability. The 'normalisation' of these symptoms considerably undermines the efforts sought to address the problem which is a major welfare concern for brachycephalic dogs.

1.4.3.2 Clinical Presentation

The pathogenesis of BOAS has long been considered a two-stage process and a downstream effect of the brachycephalic head conformation. Soft tissues within the upper airway such as the nostrils, nasal turbinates and soft palette are not thought to reduce proportionately in size with the reduced facial skeleton that houses them (Figure 1.8) (Harvey 1989). As a result of this discordance, the nostrils become stenotic (pinched together), nasal turbinates protrude caudally into the nasopharynx and the soft palate is relatively elongated and extends caudally where it can interact with the epiglottis (Figure 1.8B) (Torrez and Hunt 2006; Ginn *et al.* 2008; Bernaerts

et al. 2008; Fasanella *et al.* 2010; Cantatore *et al.* 2012; Heidenreich *et al.* 2016). These are the primary observations of BOAS which restrict and cause aberrant airflow and increase negative pressure within the airway (Harvey 1989; Torrez and Hunt 2006). Airway resistance and altered pressures are thought to then stimulate secondary pathological modification of the upper airway soft tissues. This includes the further elongation and thickening of the soft palate, eversion of the laryngeal saccules and tonsils, oedema of the nasopharynx, laryngeal collapse and tracheal hypoplasia (Koch *et al.* 2003; Torrez and Hunt 2006; White 2011). Collectively, these modifications to the airway severely hamper effective breathing and frequently cause respiratory noises (stertor/stridor), sleep apnoea and can accumulate to an intolerance to heat/exercise, cyanosis and collapse (Ginn *et al.* 2008; Bernaerts *et al.* 2008). Treatment options for BOAS are largely restricted to surgical interventions in which the soft tissues that are protruding into and blocking the airway are excised or ablated. Common procedures include the removal of the laryngeal saccules and tonsils, wedge resections of the nostrils, laser turbinectomy and excision of the causal aspect of the soft palate which have a mixed prognoses (Poncet *et al.* 2006; Torrez

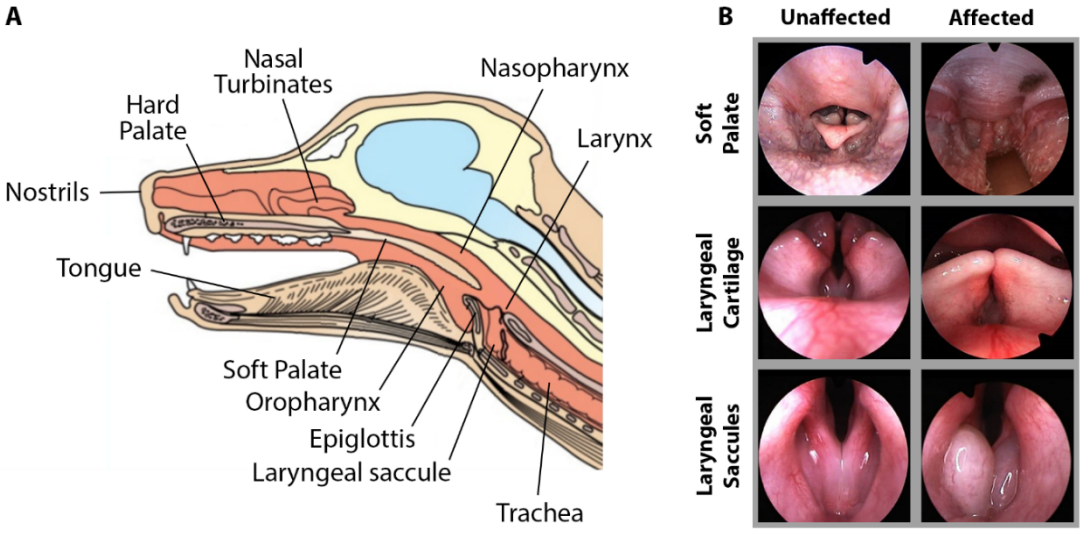


Figure 1.8 – Structures of the Upper Airway. (A) Lateral midline schematic of the dog upper airway. Diagram sourced and modified from VetHQ Online available at: http://vethq.com.au/mable/. (B) Laryngoscopic examination of Norwich Terrier unaffected (left) and affected (right) by upper airway syndrome. From top to bottom in affected dogs, the soft palate extends caudally to the epiglottis, laryngeal cartilage is inverted and laryngeal saccules are swollen.

and Hunt 2006; Riecks *et al.* 2007; Bernaerts *et al.* 2008; Fasanella *et al.* 2010; Oechtering *et al.* 2016). The extent of brachycephalic dogs requiring major upper airway surgery is set to rise further with the rapid increase in popularity of breeds such as the French bulldog in the UK, which again, highlights the major concerns for animal welfare (Kennel Club, 2018b).

1.4.3.3 Causes of BOAS

The segregation of BOAS with brachycephalic breeds is a clear indicator that there is a hereditary component to the disease (e.g. determinants of skull shape) however the genetic factor(s) are extremely limited. To date, only one variant in the *BMP3* gene has been shown to contribute to brachycephaly in a subset of small brachycephalic breeds (Schoenebeck *et al.* 2012). Beyond this, other factors such as age and weight have been inconsistently linked with the disease. Several studies have concluded that age has no bearing on the disease presentation with many dogs having clinically significant symptoms from 1 year of age and above (Poncet *et al*. 2006; Torrez and Hunt 2006; Fasanella *et al.* 2010). However, the most extreme BOAS cases with excessive secondary pathological modification of the airways are observed in dogs aged 3 years and above (Fasanella *et al.* 2010). Body condition scores were observed to be significantly higher in BOAS-diagnosed dogs and one analysis showed extreme obesity put brachycephalic breeds at high risk (Liu *et al.* 2015; 2016) Conversely, other studies concluded weight had no correlation with the disease (Torrez and Hunt 2006). The inconsistency of these findings is likely due to the relatively small (<100 dog) study cohorts, variation in diagnostic techniques and diverse study demographics across the veterinary practices conducting analyses. Long term prospective observational studies are required to answer these.

Despite this inconsistency, a frequent observation is the variability in BOAS susceptibility and heterogeneity in disease presentation across the different brachycephalic breeds (Poncet *et al.* 2005a; Torrez and Hunt 2006; Riecks *et al.* 2007; Ginn *et al.* 2008; De Lorenzi *et al.* 2009; Fasanella *et al.* 2010; Liu *et al.* 2016). Although the Bulldog, French bulldog and Pug all have similar skull conformations, five out of

six studies examining the prevalence of BOAS reported the highest incidences were in the Bulldog followed then by the French bulldog and Pug (Poncet *et al.* 2005b; Torrez and Hunt 2006; Riecks *et al.* 2007; Ginn *et al.* 2008; De Lorenzi *et al.* 2009; Fasanella *et al.* 2010). Furthermore, whole-body plethysmography revealed that even in unaffected brachycephalic breeds, the Bulldog had distinct breathing characteristics compared to the remaining brachycephalic and non-brachycephalic breeds (Liu *et al.* 2016). It is unclear whether factors such as age and weight may truly contribute to this variability – again, only larger scale prospective cohort studies could start to unravel the roles of these factors and their interactions. However, the apparent variable predisposition of BOAS across brachycephalic breeds may suggest that additional, possibly genetic and yet unknown factors also contribute to BOAS presentation.

1.4.3.4 BOAS in a Non-Brachycephalic Breed

In recent years there has been an increase in reported cases of a non-brachycephalic breed, the Norwich Terrier, suffering from symptoms remarkably similar to BOAS (Dietschi *et al.* 2010; Johnson *et al.* 2013). Within the breed the condition is called Upper Airway Syndrome (UAS) and shares many of the upper airway pathological alterations as seen in BOAS including an elongated and thickened soft palate, oedema of the pharynx and eversion of the laryngeal saccules (Dietschi *et al.* 2010; Johnson *et al.* 2013). Intriguingly, a closely related breed to the Norwich Terrier, the Norfolk Terrier is not affected by UAS suggesting there is a genetic component to the condition irrespective of the non-brachycephalic conformation of the breed.

Currently, the prevailing dogma held by many is that the extreme brachycephalic head conformation is the major predisposing factor to disease presentation. As such, one of the main research objectives of this thesis was to identify the genetic determinants of canine skull morphology.

1.5 Thesis Aims

To date, no study has explicitly addressed the genetic causes of Brachycephalic Obstructive Airway Syndrome in the dog. As highlighted previously, studies have catalogued the pathological changes in the airway, the prevalence of the condition and novel approaches to grading its severity however its cause remains elusive. The current consensus within the field is that the brachycephalic head conformation is the major predisposing factor to the condition which itself, also has a restricted genetic understanding.

Geometric morphometric analyses of human craniofacial variation were successful in isolating multiple positional candidate genes linked with craniofacial diversity. This thesis will exploit a similar method of identifying positional gene candidates albeit with high-resolution computer tomography scans of the skull rather than stereophotogrammetric images of superficial soft tissues of the face. Importantly, the biological relevance of variants associated with canine brachycephaly will be explored before directly addressing the presentation of respiratory distress in dogs.

The major research objectives of this thesis can be summarised as follows:

1. Develop a method to quantify canine craniofacial shapes and best account for the confounding influences of allometry.
2. Identify genomic loci driving craniofacial morphology in the dog and find causal variants within them.
3. Characterise the molecular and phenotypic effects of variants within the *SMOC2* gene associated with canine brachycephaly.
4. Use a non-brachycephalic dog breed to find genes that cause respiratory distress across dogs irrespective of their craniofacial conformation.

Chapter 2: Materials and Methods

2.1 gDNA Extraction from Whole Blood

Following written consent from owners, residual diagnostic whole blood in EDTA-containing 2 mL vials was stored at 4 °C by clinicians at the HfSA. Samples were transferred to the Roslin Institute for extraction.

2.1.1 Preparation of Stock Extraction Buffers

Stocks of Extraction Buffer A and Extraction Buffer B were prepared in advance. Buffer A (red blood cell lysis buffer) stock was made by adding 109.54 g 0.32 M sucrose and 1.02 g 5 mM hydrated magnesium chloride to 10 mL 1 mM Tri-HCl in a 1 L glass bottle. The mixture was made up to 950 mL with Milli-Q® (ZRXQ003WW, Merck) water. The pH was adjusted to 7.6 with 0.2 M hydrochloric acid (12920934, Fisher Scientific) (see Section 2.1.2). The pH adjusted solution was vacuum filtered by pouring into the top chamber of a Nalgene™ Rapid-Flow™ filter unit (127-0020, Thermo Scientific™) and applying a steady flow to the tap. Once complete, the solution was kept in the 1 L plastic flow-through bottle supplied as part of the filter kit and 7.5 mL 0.75% Triton™ X-100 (10102913, Fisher Scientific) was added. The final buffer volume was made to 1 L with Milli-Q® water.

Buffer B stock was made by mixing 20 mL 1 M Tris-HCl, 8 mL 0.5 M disodium EDTA and 20 mL 5 M sodium chloride in a 1 L glass bottle which was made up to 950 mL with Milli-Q® water. The pH was adjusted to 7.4 with 0.2 M hydrochloric acid and filter sterilised as described above. The filtered solution was kept in the 1 L plastic flow-through bottle and made up to 1 L with Milli-Q® water.

2.1.2 Adjusting the pH of a Solution

A magnetic stirrer bar (11507812, Fisher Scientific) was cleaned in 70% ethanol – 7 parts 100% ethanol (10342652, Fisher Scientific): 3 parts Milli-Q® water. The bar was placed in the solution to be pH adjusted. The bottle was placed on the magnetic

stirrer plate (CB161, Bibby Scientific) and stirred at room temperature at a rate to induce a small vortex. The S20 SevenEasy™ pH meter (1231165663, Mettler Toledo) was turned on and the probe was removed from the storage bottle. A wash bottle containing deionised water was used to rinse the probe above a waste jar. To calibrate, the probe was placed in the pH 10 solution until the monitor stopped flashing. The probe was rinsed with deionised water and placed in the pH 7 solution. The probe was subsequently rinsed, placed in the pH 4 solution and rinsed once more to complete the calibration. Excess water was removed from the probe with a lint-free medical wipe (7432, Kimberly-Clark Professional) before using the mounting bracket to hold the probe in the solution on the magnetic stirrer. To decrease the pH, ~1 mL 0.2 M hydrochloric acid (12920934, Fisher Scientific) was added at a time and the 'Read' button was pressed on the monitor to ensure measurements are up to date. The hydrochloric acid was not added directly to the pH probe. Once the pH was reduced to the correct level, the probe was rinsed in deionised water and returned to the storage buffer. The pH meter was turned off.

2.1.3 Extraction Protocol

Stock Buffer A solution was mixed in a 1:2 ratio with deionised water to create a Buffer A working solution. Whole blood residual samples were transferred to a 15 mL Falcon™ tube (188271, Greiner Bio-One) with a 1 mL graduated Pastette (LW4040, Alpha Laboratories Ltd.). Buffer A working solution was used to clean and transfer samples to the Falcon™ tube. Buffer A working solution was added to a final volume of ~5 mL. Samples were gently inverted 10 times and centrifuged (5810 R, Eppendorf) at 4,000 rpm for 5 minutes before decanting the supernatant in a waste bottle containing 2 effervescent haz-tabs (H8811, Guest Medical).

Stock Buffer A was mixed in a 1:3 ratio with deionised water and 5ml used to re-suspend the cell pellet before centrifuging at 4,000 rpm for 5 minutes and decanting the supernatant. Stock Buffer B was mixed in a 10:1 ratio with 10% SDS (151-21-3, Sigma-Aldrich) and 2.5 mL used to re-suspend the pellet before 25 μL 20mg/mL proteinase K (AM2546, ThermoFisher Scientific) was added. Samples were

shaken overnight in an air incubator (Innova™ 4000, New Brunswick Scientific) at 55 °C before 2.5 mL 5 M sodium chloride solution was added. Samples were inverted 10 times before centrifuging at 4,000 rpm for 20 minutes. The supernatant was transferred to a new 15 mL Falcon™ tube containing 5 mL 100% isopropanol (BP26181, Fisher Scientific). A plastic Pastette was used to ensure the complete transfer of the supernatant without disturbing the pellet. The supernatant and isopropanol were mixed by inversion and air incubated at room temperature overnight on the benchtop. Samples were centrifuged at 4,000 rpm for 45 minutes at 4 °C before the supernatant was decanted and the pellet was gently washed with 1 mL 70% ethanol for 2 minutes. Samples were centrifuged at 4,000 rpm for 10 minutes before the supernatant was decanted and the pellet was air dried overnight. gDNA was hydrated by adding 200 μL 1X TE and allowed to re-suspend at room temperature overnight. Samples were transferred to labelled 1.5 mL screw-cap microfuge tubes and stored at 4 °C.

2.2 gDNA Extraction from Tissue Biopsy

Following written consent from owners, excised tissue removed in the course of planned surgical treatments was stored in 7 mL bijou tubes at -20 °C by clinicians at the HfSA. Samples were transferred to the Roslin Institute for extraction. All samples were extracted using the same protocol regardless of tissue type.

An extraction buffer was prepared by mixing 5 mL 1 M Tris (pH 8.0), 10 mL 0.5 M EDTA (pH 8.0), 25 mL 20% SDS, 10 mL 5 M sodium chloride solution and 450 mL deionised water. Frozen tissue samples were cut into~0.5 cm^3 pieces on lint-free paper (DR670, Berkshire) using a scalpel (P105/5, Paragon®) and forceps (F6521, Sigma-Aldrich). The tissue was allowed to thaw. For each extraction, one ~0.5 cm^3 sample was used and gently cut into smaller pieces before adding to 750 μL extraction buffer, 40 μL 1 M DTT (3483-12-3, Sigma-Aldrich) and 15 μL 20mg/mL proteinase K solution (Thermo Fisher Scientific, 25530049) in a 1.5 mL screw-top microfuge tube. Samples were incubated 60 °C overnight in a Dri-Block® heat block (DB.2A, Techne) until complete digestion was achieved. To assist gDNA precipitation,

270 μL saturated sodium chloride solution was added to the digested samples and vortexed using a Vortex-Genie™ 2 (15547335, Fisher Scientific) for 20 seconds. Samples were centrifuged at 15,000 G (5415D, Eppendorf) for 15 minutes before the supernatant was transferred to a new 1.5 mL screw-cap microfuge tube. gDNA was precipitated using 100% ethanol and pelleted in a centrifuge at 15,000 G for 15 minutes. The supernatant was decanted and the pellet was gently washed with 200 μL 70% ethanol for 2 minutes before being centrifuged for a further 15 minutes at 15,000 G. The supernatant was decanted and the pellet was allowed to air-dry overnight inverted on lint-free paper. gDNA was hydrated by adding 200 μL TE and allowed to re-suspend at room temperature overnight. Samples were transferred to labelled 1.5 mL screw-cap microfuge tubes and stored at 4 °C.

2.3 gDNA Extraction from Buccal Swab

The PERFORMAgene™ buccal swab (DNA Genotek, PG-100) is a single-use kit designed to collect, stabilise and transport sources of animal gDNA. The kit is comprised of an exposed sponge mounted on a screw cap attached to a tube. The sponge is placed in the cheek pouch of the dog and gently moved for up to 30 seconds to collect the pooling saliva. Once collected, the kit is held upright, the cap-mounted sponge is unscrewed and inverted into the tube and resealed tightly. The tube contains a stabilising storage buffer. Sealed kits should be shaken vigorously 10 times to ensure the sample and storage buffer are mixed. The kit is stored at room temperature until extraction.

With written consent from the owners, the PERFORMAgene™ buccal swabs were used by clinicians at the HfSA for patients which did not have available residual diagnostic whole blood or excised tissues following surgery. The buccal swabs were also used as part of the patient outreach initiative.

Prior to gDNA extraction, samples were vigorously mixed by inversion for 5 seconds and placed in an air incubator at 50 °C overnight. The collection sponge was pressed and rotated against the inside of a 15 mL Falcon™ tube to maximise gDNA

recovery. Samples were not poured to avoid the transfer of any debris in the collection tube. The sample, totalling ~500 μL was transferred from the Falcon™ tube to a 1.5 mL screw-top microfuge tube which contained 20 μL purifier solution (DNA Genotek, PG-L2P). Samples were vortexed for 2 seconds and incubated on ice for 10 minutes. After samples were centrifuged at 15,000 G for 5 minutes, the supernatant was carefully transferred to a fresh 1.5 mL screw-top microfuge tube without disturbing the pellet. To assist gDNA precipitation, 25 μL 5 M sodium chloride solution was added and mixed by inversion. gDNA was precipitated by adding 600 μL 100% ethanol and mixing by inverting 10 times. To ensure complete precipitation, samples were incubated at room temperature overnight on the benchtop. gDNA was pelleted by centrifuging at 15,000 G for 2 minutes prior to carefully decanting the supernatant. The gDNA pellet was washed gently in 250 μL of 70% ethanol for 2 minutes at room temperature. Samples were centrifuged again at 15,000 G for 2 minutes and allowed to air dry overnight by leaving the microfuge tubes open on the benchtop. gDNA was hydrated by adding 200 μL TE and allowed to re-suspend at room temperature overnight. Samples were transferred to labelled 1.5 mL screw-cap microfuge tubes and stored at 4 °C.

2.4 gDNA Quantification and Quality Assessment

2.4.1 NanoDrop Device

Rough indications of gDNA concentration and quality were quantified using a NanoDrop 1000 spectrophotometer (Thermo Scientific). The 'Nucleic Acid' tab in the NanoDrop software was used to quantify gDNA. The software requests that the NanoDrop device is prepared by lifting the spectrophotometer arm, pipetting 1 μL deionised water onto the measurement pedestal, gently lowering the arm and clicking 'OK'. Before and between loading samples, the arm and pedestal are wiped clean with a lint-free tissue. Subsequently, 1 μL TE was added to the pedestal and measured by clicking 'Calibrate'. To measure gDNA samples, the name of the sample was typed input and 1 μL was loaded onto the pedestal and quantified by clicking

'Run'. All measurements were exported and saved as a plain text file. Samples were considered to be pure with a 260:280 nm absorbance ratio of ~1.8.

2.4.2 Qubit™ dsDNA Broad Range Assay Kit

The concentrations of gDNA samples destined for whole genome sequencing were accurately quantified using the Qubit™ instrument (small sample sizes). For gDNA quantification within the range of 100 pg/μL – 1,000 ng/μL the Qubit fluorometric device (v4) was used with the Qubit™ dsDNA BR assay kit (ThermoFisher Scientific, Q32850). One 0.5 mL flip-cap microfuge tube (Q32856, Life Technologies) was labelled for Standard 1, Standard 2 and for each sample to be assayed. A working solution was prepared in a 7 mL plastic bijou tube (129A, Thermo Scientific) by diluting the Qubit® dsDNA BR Reagent 1:200 in Qubit® dsDNA BR buffer. The final volume of working solution accommodated for 200 μL per microfuge tube prepared previously. To prepare the standards, 10 μL of the Qubit® dsDNA BR Standard #1 was added to 190 μL working solution in the labelled microfuge tube. The second standard was similarly prepared using the Qubit® dsDNA BR Standard #2. Both standards were mixed by vortex for 2 seconds. To prepare the samples, 1 μL of each sample was added to 199 μL working solution in the appropriately labelled microfuge tubes and mixed by vortex for 2 seconds. Samples were incubated for 2 minutes at room temperature. All steps were carried out with minimal exposure to light. The Qubit device was turned on and the 'dsDNA Broad Range' tab was selected. The input sample volume was selected as '1 μL'. Standard 1 was placed in the device beneath the arm and measured by pressing 'Run Standard 1'. This step was repeated for standard 2. Once calibrated, the samples were measured in turn by placing in the device and clicking 'Measure Sample'. gDNA concentrations were recorded.

2.4.3 PicoGreen™ Assay

The concentrations of large batches of gDNA samples destined for SNP microarray genotyping were accurately quantified using the Quant-iT™ PicoGreen® dsDNA quantification kit (ThermoFisher Scientific, P11496).

2.4.3.1 Standard Curve Preparation

A serial dilution of the 100 µg/mL bacteriophage lambda (λ) DNA was used to create a DNA standard curve. To do this, eight 0.5 mL microfuge tubes were labelled 1-8. To tube 1, 490 µL TE and 10 µL 100 µg/mL bacteriophage lambda (λ) DNA were added before mixing by vortex for 5 seconds. Tubes 2-8 all contained 250 µL TE. After vortexing, 250 µL Tube 1 was diluted into Tube 2. Tube 2 was then vortexed before transferring 250 µL to Tube 3. The serial dilution was completed for all tubes following the instructions in Table 2.1. All samples were vortexed between dilution steps.

2.4.3.2 DNA Sample Preparation

To a 96-well PCR plate, 199 µL TE and 1 µL gDNA sample were added. Plates were sealed with an adhesive plastic cover (AB-0558, Thermo Scientific) before being mixed by inversion and collected at the bottom of the plate by briefly centrifuging.

2.4.3.3 Working Solution and Plate Preparation

The total volume of required PicoGreen® Working Solution is determined by adding 21 (for standards) to the number of gDNA samples and then multiplying by 25 µL. The total volume of PicoGreen® Working Solution is made up with a 1:200 dilution of Quant-iT™ PicoGreen® dsDNA reagent to TE before mixing by vortex. The PicoGreen® Working Solution is prepared in a 7 mL bijou tube wrapped in tin foil to protect from photo-bleaching. To a flat bottomed 96-well plate (167008, Thermo Scientific), 25 µL of bacteriophage λ DNA standards were loaded into the wells listed in Table 2.1. The remaining wells were loaded with 25 µL of each diluted gDNA sample. The PicoGreen® Working Solution was poured into a multichannel trough (12369175, Fisher Scientific) and 25 µL was transferred to every well of the flat-bottomed plate using a multichannel pipette. The trough was covered with tin foil to prevent photo-bleaching. The plate was gently swirled to mix.

2.4.3.4 Plate Reading and Analysis

The Perkin Elmer Wallac 1420 VICTOR2™ plate reader was turned on and the flat-bottomed 96-well plate was loaded into the plate compartment with well A1 being orientated to the top-left corner. The Wallac software was loaded and the 'Fluorescein High Count (1s)' option was selected from the Choose Program' tab. The 'START' button was clicked to allow the scanner to run for a duration of ~5 minutes. The 'Results of the Latest Assay Run' option was selected from the 'TOOLS' menu and a text file containing the results was exported by selecting 'Export' from the 'FILE' menu. The Wallac software and device was exited and turned off after use.

The results from the Wallac were opened in Microsoft Excel. The Fluorescein (1s) counts of the bacteriophage λ DNA serial dilutions were plotted against their concentration in pg/µL as listed in Table 2.1 to create a standard curve. The concentration of the gDNA samples were predicted from their Fluorescein counts using the standard curve and the "=TREND" function in Microsoft Excel. When using the "=TREND" function, the concentrations of the standard dilutions were selected as "known_y's", the Fluorescein (1s) counts of the standard dilutions were selected as "kown_x's" and the individual gDNA sample Fluorescein count was selected for the "new_x's". The "const" factor was set as 'TRUE'. This was applied to all gDNA sample Fluorescein counts. The final gDNA concentration in ng/µL was calculated by multiplying the results by 0.4 to account for the dilution during preparation.

Tube	Concentration (pg/µL)	Volumes	Add to Wells
1	1,000	10 µL λ DNA Standard, 490 µL TE	-
2	500	250 µL Tube 1, 250 µL TE	A1 - A3
3	250	250 µL Tube 2, 250 µL TE	B1 - B3
4	125	250 µL Tube 3, 250 µL TE	C1 - C3
5	62.5	250 µL Tube 4, 250 µL TE	D1 - D3
6	31.25	250 µL Tube 5, 250 µL TE	E1 - E3
7	16.625	250 µL Tube 6, 250 µL TE	F1 - F3
8	0	250 µL TE	G1 - G3

Table 2.1 - PicoGreen Assay Preparation. Volumes of DNA standards required to produce the appropriate concentrations for the assay.

2.4.4 TapeStation Device

The quality of gDNA samples destined for whole genome sequencing were calculated using TapeStation 2200 (Agilent Technologies), Gnomic DNA ScreenTape (Agilent Technologies, 5067-5365) and Genomic DNA Reagents (Agilent Technologies, 5067-5366).

Genomic DNA Reagents were allowed to equilibrate to room temperature for 30 minutes before vortexing for use. A genomic ladder solution was prepared by mixing 10 μL Genomic DNA Sample Buffer with 1 μL Genomic DNA ladder in the first well of an Optical Tube 8-well strip (Agilent Technologies, 5067- 5150). The gDNA samples were prepared by mixing 1 μL gDNA sample with 10 μL Genomic DNA Sample Buffer. All samples were collected at the bottom of the tube by briefly spinning in a strip tube centrifuge (PMC-860, Stratagene) before being vortexed for 1 minute at 2000 rpm. Tubes were spun down before being inserted into the TapeStation instrument. To prepare the TapeStation, fresh loading tips (Agilent Technologies, 5067- 5152) were inserted into the left-hand side of the tip cassette and the Genomic DNA ScreenTape was slotted into its holder such that the QR code could be read by the scanner on the right of the device. The TapeStation software was opened and the wells containing the ladder and samples were selected. The 'START' button was clicked and the results were exported as a PDF file once finished. Once complete, used loading tips were discarded and fresh tips were loaded. A DNA integrity number (DIN) of ~8.0 and above is considered to represent good quality gDNA.

2.5 Polymerase Chain Reaction (PCR)

2.5.1 Standard KOD Xtreme™ Hot Start DNA Polymerase Reaction Mix

The TProfessional Gradient thermocycler (Biometra®) was used for all PCR experiments with the KOD Xtreme™ Hot Start DNA Polymerase kit (Merck, 71975-3). The kit included KOD Xtreme™ DNA polymerase, 2X Xtreme™ buffer and 2 mM dNTPs. Thermocycler conditions and primers vary between PCRs however a standardized reaction mixture was used for all protocols (Table 2.2).

Component	Volume for 1x Reaction (µL)
2X Xtreme™ Buffer	5.0
dNTPs (2mM)	2.0
Deionized Water	0.5
Forward & Reverse Primers (10 µM)	0.3
DNA (5 µg/µL)	2.0
KOD Xtreme™ Hot Start DNA Polymerase	0.2

Table 2.2 – KOD Xtreme™ Reaction Mixture. The components and volumes of a standard KOD Xtreme™ reaction mixture.

2.5.2 Standard Thermocycler Programme

A commonly used template for thermocycler settings for PCRs is outlined in Table 2.3. The temperature for the primer annealing step is dependent on the specific primers used. Primer annealing temperatures were predicted using the Primer3 online software (see Section 2.5.4). The extension time was also variable and was dependent on the size of product being amplified. A guideline extension time of 1 min/kb is given for the KOD Xtreme™ Hot Start DNA Polymerase. For example, for a predicted 1,200 bp product, an extension time of 90 seconds would be used to ensure complete extension.

Cycles	Temperature (°C)	Time
1x	95	2 minutes
40x	98	15 seconds
	variable	30 seconds
	68	variable
1x	72	7 minutes

Table 2.3 – Standard Thermocycler Programme. The temperatures and duration of steps in a standard thermocycler programme. The temperature of the annealing step and duration of the extension phase are reaction-dependent.

2.5.3 Touchdown Thermocycler Programme

Some amplicons require the use of a touchdown thermocycler programme which is split into two sections. The first section starts with an annealing temperature

approximately 10 °C above the optimal temperature (as predicated by Primer3) and is successively reduced by 1 °C for every cycle. This is repeated up to 15 times and is intended to exclusively select the target amplicon and reduce off-target amplification. The second phase then runs 25 cycles at the optimal annealing temperature to preferentially use the template generated in the previous cycles. Extension times are calculated by 1 min/Kb.

Cycles	Temperature (°C)	Time
1x	95	2 min
15x	98	10 s
	Annealing temp - 1 °C/cycle	30 s
	72	variable
25x	98	10 s
	variable	30 s
	72	variable
1x	72	7 min

Table 2.4 – Touchdown Thermocycler Programme. The temperatures and duration of steps in a touchdown thermocycler programme. The temperature of the annealing step and duration of the extension phase are reaction-dependent.

2.5.4 Primer Design

gDNA sequence surrounding the target amplicon was inserted into the Primer3 (v4.1.0) web input page (http://primer3.ut.ee) (Untergasser *et al.* 2012). Forward and reverse primers were selected using the following parameters: primer size 18-23 bp, primer temperature 50-70 °C, GC content 30-70 % with Breslauer et al. 1986 thermocycler parameters. Optimal annealing temperatures for primers were recorded and primer sequences were pasted into the custom oligo ordering system from Integrated DNA Technologies (https://www.idtdna.com/site/order/oligoentry). The '25 nM DNA oligo' scale and a 'standard desalting' purification process was selected from the dropdown menus. Dried oligos were resuspended in 1x TE to achieve a stock concentration of 100 µM which were stored at -20 °C. Working primer concentrations of 10 µM primers were created by diluting in deionised water.

2.5.5 Primer Sequences

Primer sequences designed previously to amplify and genotype variants associated with canine morphological variation were acquired from the relevant published literature (Table 2.5) (Parker *et al.* 2009; Schoenebeck *et al.* 2012; Rimbault *et al.* 2013; Marchant *et al.* 2017). Primer sequences were also designed here to amplify specific variants in this thesis (Table 2.5).

2.5.6 Genotyping

2.5.6.1 Preparation of Master Plates

In order to genotype up to nine morphology-associated QTL for all study participants, master working plates of gDNA dilutions were prepared. A final concentration of 5 ng/µL was prepared for each sample in a total volume of 60 µL. To achieve this, varying volumes of gDNA were added to 96-well plates to ensure each well had a total of 300 ng gDNA. The plates were dried using a speed vac (SC250EXP, Thermo Scientific) for 15 minutes. gDNA samples were re-hydrated in 60 µL deionised water and allowed to resuspend over night at room temperate.

2.5.6.2 Structural Variants

Structural variants (*SMOC2*, *IGF1*, *SMAD2* and the *FGF4* retrogene) were genotyped based on the amplicon size visualised by 2% agarose gels. To prepare a gel, 1 L 1X TAE buffer was created by mixing 20 mL 50X TAE buffer with 980 mL Milli-Q® water. To create a 100 mL gel, 2 g Ultrapure™ Agarose (16500500, Thermo Scientific) was mixed with 100 mL 1X TAE in a 250 mL glass bottle and heated in a microwave (NN-E442W, Panasonic) for ~2-3 until completely dissolved. Once the bottle has cooled for 2 minutes, 15 µL 10,000X SYBR Safe DNA gel stain (Invitrogen, S33102) was added and mixed by swirling. The gel was cast in a gel dock (Bio-Rad) with combs and allowed to set for 2 hours under tin foil. Once set, the gel was place in a Sub-Cell® electrophoresis tank (170-4401, Bio-Rad) and submerged in 1X TAE. Following the PCR, 2 µL PCR products were added to wells of an 8-well strip-tube which contained 8 µL deionised water and 2 µL blue/orange 6X loading buffer

(G190A, Promega). The samples were mixed briefly by vortex and collected at the bottom of the tube by centrifuging for 10 seconds. Mixes were loaded into the prepared agarose gel and ran at 100 V for 30 minutes using the Bio-Rad PowerPak™ Basic system. A 1 kb Plus DNA ladder (10787018, Fisher Scientific) was added into the first and last lanes. Gels were visualised using a Gel Logic 200 Imaging System (Kodak). Gels were placed on the bed of the imager. The Kodak 1D software (v3.6.5 K2) was loaded and a 'New Image Capture' session was selected. The image of the gel was magnified and focussed using the white light live settings before switching to the fluorescent transmission setting to capture the banding patterns of the gel. Images were saved and printed to allow for banding interpretation and genotyping.

2.5.6.3 SNPs and small INDELs

Amplicons targeted at SNP and small INDEL variants (*BMP3*, *ITGA11*, *IGF1R*, *GHR*(1), *GHR*(2) and *ADAMTS3*) were sanger sequenced by either Edinburgh Genomics, UK or The Institute of Genetics & Molecular Medicine, Edinburgh, UK. Samples in 96-well plates or 8-well strip-tubes were delivered in person with an excess (2 µL required per sample) of 10 µM forward and reverse primers. Samples were given as 'dirty PCR products' and required no prior cleaning steps. Upon the receipt of raw sequence and sequence traces, reads were entered and aligned using the 'Pro' setting in SeqMan Pro™ (v10.1.0). Genotype calls were made by a single analyst.

2.6 Generating Plots and Figures

Throughout this thesis, R (v3.3.0) and RStudio (v1.0.153) have been used to generate graphics and to analyse data. Specific R packages are stated when used but common packages include: ggplot2, stringr, reshape2, plyr, etc. Figures and data analysis out with this programme are explicitly cited throughout but may include programmes such as Microsoft Excel 2013, ImageJ (Schneider *et al.* 2012) and MorphoJ (Klingenberg 2010) for analysis and Adobe Photoshop CC 2015 for figure construction.

Locus (position)	Forward Primer	Conditions (temperature, time)	Ref
IGF1 (15:41,220,980)	CACTGATCCAGAAGAATCCAACT (F)	62.4 °C, 1 min	Rim
	CAAAGAACCATGTAAGCCTATTTGT (R)		
IGF1R (3:41,849,479)	AGATGACCAACCTCAAGGATATT (F)	61 °C, 1 min	Rim
	AGTCCTGCCATCCCACAAAG (R)		
GHR (1) 4:67,040,898 (2) 4:67,040,939	GCTCTCCGTTAAATCAAGCTG (F)	56 °C, 45 s	Rim
	AAGGAGAGAGGTGTTGTTGGT (R)		
STC2 (4:39,182,836)	ATACAATCCACCTAGTGTCCCCAACCAT (F)	60 °C, 30 s	Rim
	GGCCACAGCCCCTTTAAT (R)		
BMP3 (32:5,231,894)	GATACAGGAGATTGTGCCAAATGGGTAA (F)	67.6 °C, 30 s	Sch
	CTCCTGGTGGAAATCGTCAGTCTATCTG (R)		
SMAD2 (7:43,794,129)	GCTTCAAGTCAGTGTGCTCC (F)	55 °C, 1 min	Mar
	CGTATTTGTTGCTGCTGGGT (RA)		
	AGAGCCCTGACATCATGACC (RD)		
FGF4 retrogene (18:20,343,712-20,443,725)	CACACAGATGGACCATGAAA (F)	68 °C, 1 min	Mar
	TTTTAGATTCCGCACATGAG (RA)		
	CTCTTTGAACTTGCACTCCTC (RD)		
ITGA11 (30:32,354,583-32,354,585)	AGGGATAGTCCAGCTCCAAGGCTGGTAT (F)	70 °C, 30 s	Mar
	CTCTTTCAGGCTTCCCCAGTTGTACCTA (R)		
SMOC2 (1:55,939,775-55,941,305)	GGCAGGGGATGGGGAAGGCT (F)	72-52 °C, 55 °C, 45 s *	Mar
	ACTGTGTGCTTTGCCCAAACTCA (RA)		
	TGCCCATAAAGTTCAGGGTCCACT (RD)		
ADAMTS3 (N/A)	ACACACGAACCCAGGCACAC (F)	62.4 °C, 1 min	Mar
	GGCCTGGGAGCACTGCAC (R)		

Table 2.5 - PCR Primers. The positional gene candidate and its reported variant(s) positions. Primers are denoted as forward (F), reverse (R) ancestral (RA) and reverse derived (RD). All PCR reactions use standard thermocycler conditions with primer annealing temperatures and extension times given respectively in the conditions column. *denotes the use of a touchdown thermocycler programme. Primer sequences are taken from (Rimbault *et al.* 2013) (Rim), (Schoenebeck *et al.* 2012) (Sch) and (Marchant *et al.* 2017) (Mar).

Chapter 3: Capturing Morphological Variation in the Domestic Dog Skull

3.1 Introduction

Many morphological traits of the domestic dog have changed significantly since its derivation from its ancestor, the grey wolf. The dog now occupies a morphospace which is distinct and far exceeds that of the wolf (Drake *et al.* 2015; 2017). This morphological distinction is hypothesised to have at least partly been a direct response to the domestication process. Fossilised remains of wolf-like dog progenitors have been found across Europe from similar time points to the predicted onset of domestication based on molecular DNA-based clocks (Tchernov and Valla 1997; Savolainen *et al.* 2002; Germonpré *et al.* 2009; Ovodov *et al.* 2011; Pionnier-Capitan *et al.* 2011; Germonpré *et al.* 2012; Thalmann *et al.* 2013; Freedman *et al.* 2014; Frantz *et al.* 2016), indicating that morphological changes were observed early in domestication. These observations were supported by a controlled domestication experiment of a taxonomically close relative of the dog, the silver fox (Trut *et al.* 2009). Over a period of 40 years, wild silver foxes were selectively bred for 'tamability' alone and as a result the foxes underwent some profound morphological changes. Tame foxes lost pigmentation across their coats, their ears drooped, legs and tails became shorter and they developed under- or overbites. These changes were only observed after 15-20 generations in a small number of foxes however similar morphological changes have been seen across domestic animals suggesting that genetic and morphological changes occur in response to selection for tameness during domestication (Zeder 2012).

The eighteenth and nineteenth centuries were the next major landmark in the morphological variability of the dog. During this time, breeding became popular and coincided with the establishment of the breed clubs (Wayne and Ostrander 2007). Breeders quickly recognised and exploited the ease of propagating desirable morphological traits within the closed populations of breed boundaries. In many

instances, the accentuated morphological features were restricted to specific breeds and groups of breeds which then became breed-defining and a requirement in breed standards (Kennel Club 2018a). The hypervariability amongst some breeds of dog are in fact comparable to the morphological variation across some genera of wild canids (Wayne 1986). Some morphological traits continue to vary and become more accentuated over time reflecting the whimsical desires of breeders (Drake and Klingenberg 2007).

Common variable traits across the dog include body size, leg length, various coat characteristics and arguably one of the most morphologically diverse features, the canine skull (Stockard 1941). Variation in the anatomy of the dog skull has fascinated biologists for many years with the most striking craniofacial variability within the width and length of the skull. Canine brachycephaly represents an extreme skull conformation with morphological parallels to the human trait from which it takes its name (Persing *et al.* 1989). Bones of the viscerocranium including the nasal, incisive and maxillary bones have a reduced rostrocaudal length and a premature fusion of the cranial base/chondrocranium give the characteristic short and broad skulls (Huber 1974; Nussbaumer 1978). The popularity of the brachycephalic conformation was thought to be due to their apparent child-like (paedomorphic) appearance with relatively enlarged cranium, broad face and wideset eyes (Wayne 1986; Coppinger and Coppinger 2001; Drake and Klingenberg 2010; Drake 2011). The retention of the juvenile features in the adult dog was thought to simply be due to an arrestment of the normal wolf skull development at an infant stage (heterochrony) however, recent work suggests it is in fact neomorphic (Drake 2011).

In addition to the brachycephalic versus dolichocephalic ("long-faced") craniofacial variation, notable changes in the prebasial angle are observed across the domestic dog. The prebasial angle is the angle between the plane of the hard palate and the plane of the cranial base with the extreme klinorhynchy and airorhynchy describing the ventral and dorsal rotation of the palate relative to the cranial base, respectively (Drake and Klingenberg 2010; Mitchell *et al.* 2014; Geiger and Haussman

2016; Knowler *et al.* 2017). The extremes of klinorhynchy is most notably observed in the Bull Terrier with its characteristic "downward-tilted" foreface whilst many of the brachycephalic dog breeds have an "upward-tilted" palate (airorhynchy). Similarly to the detrimental effects of brachycephaly on health, variability in the prebasial angle is also associated with morbidities in the dog. To date, assessments have focussed on the correlation of airorhynchy with neurological defects such as syringomyelia and Chiari-like malformations however, it is unclear to what extent the prebasial angle may have on disease presentation due to the co-segregation of the brachycephalic skull conformation with airorhynchy (Mitchell *et al.* 2014; Knowler *et al.* 2017).

Finally, distinct craniofacial variation is also observed in the gross size of the dog skull. This characteristic is reflected in the more widely studied body size variation which can vary over 40-fold from <2.5 kg in the Chihuahua to >100 kg in the Mastiff (Kennel Club 2018a). Similarly, the Chihuahua is also one of the shortest breeds which measures at <20 cm at the height of the withers compared to >80 cm in the Great Dane (Kennel Club 2018a). This wide variation in the dog has been exploited to model and infer genetic contributions to adult weight and height in humans (Boyko *et al.* 2010; Hayward *et al.* 2016; Plassais *et al.* 2017). In the context of this research, the extreme variability in cranial size introduces the potentially confounding effects of allometry (see Section 1.4.2.1). The non-linear scaling of canine craniofacial features with the diverse variation in the size of the skull must be accounted for when studying purely craniofacial shape alone.

To date, four studies have directly assessed the morphological variation of the canine skull with the intention of determining the underlying genetic contributors to canine brachycephaly – the genetic results of which will be discussed in Chapter 4 (Jones *et al.* 2008; Bannasch *et al.* 2010; Boyko *et al.* 2010; Schoenebeck *et al.* 2012). Each of these studies adopted different techniques to evaluating anatomical changes in the canine skull. Jones et al., 2008 took linear measurements of the skull and snout from photographs to generate 145 breed average ratios meanwhile Bannasch et al.,

2010 compared 'classic' brachycephalic against non-brachycephalic breeds (Jones *et al.* 2008; Bannasch *et al.* 2010). Later, Boyko et al., 2010 used breed average linear skull measurements from skeletal remains whilst Schoenebeck et al., 2012 adopted the use of a MicroScribe Digitiser to capture 3D geometric coordinates of the skull (Boyko *et al.* 2010; Schoenebeck *et al.* 2012). The MicroScribe Digitiser is a mechanical arm which records the geometric position of landmarks across the ventral and dorsal surfaces of the skull before combining them *in silico*. The use of 3D coordinates to asses anatomical variation across individuals is known as geometric morphometrics and has proven to be a powerful approach to comparing the morphologies of the canine skull (Schoenebeck *et al.* 2012).

Geometric morphometrics is a relatively recent approach (within the past two decades) of assessing the overall shape of an object based on a coordinate-based system. This powerful technique can distinguish subtle changes in shape between individuals (Drake *et al.* 2015; 2017) and has largely superseded the traditional morphometric approach which used length, width and angle measurements alone (Rohlf and Marcus 1993). An appealing characteristic of geometric morphometrics is the ability to recreate objects from their coordinates and to give an intuitive and a more easily interpretable representation of the precise changes both within regions of an individual sample and across subjects (Webster and Sheets 2010). Furthermore, confounding influences on shape change are more easily isolated and removed from an object by using one of several methods of superimposition. A Procrustes superimposition is one of the most common and it allows for the removal of potential errors derived from the location, orientation and scale of the original object (Mitteroecker and Gunz 2009; Webster and Sheets 2010). Next, the centroid size of an object correlates strongly with its overall size and so has been used previously to regress the influence of allometry on the shape of an object (Webster and Sheets 2010; Drake 2011; Schoenebeck *et al.* 2012). The centroid of a 3D object is the mean, central position of all the landmarks and the centroid size is a summed measure of the distance from the centroid to each of its constituent landmarks (Bookstein 1992). Finally, dimension reduction techniques are needed to simplify the complex pattern

of variability across an object, particularly if a linear scoring system for morphological variation is needed e.g. for a downstream GWAS input. A commonly used method is a principal component analysis (PCA) which was originally described by Pearson in 1901 (Pearson 1901). PCA is a means to extract patterns and trends from high dimensionality data by aiming to explain the greatest possible proportion of correlated variation within a dataset.

The aforementioned studies assessing the morphological characteristics of the canine skull have exclusively focussed on pedigree breeds (Jones *et al.* 2008; Bannasch *et al.* 2010; Boyko *et al.* 2010; Schoenebeck *et al.* 2012). In doing so, the studies have been able to draw from the assumption that all dogs of the same breed are morphological homogeneous which permitted for the inclusion of more study participants that may not necessarily have both a phenotype and genotype. A drawback of this approach has been the oversight of potential variation within the confines of the breed boundaries and the neglecting of mixed-breed dogs. The mixed-breed population is predicted to represent as many as one third of UK dogs (3 million dogs) which means the inclusion of this population in a study would not only be a significant increment in potential study participants, but also represent a highly variable population, possibly with unique combinations of morphological traits (PFMA 2018). Furthermore, downstream genomic assessment of genetic contributions to morphological variation can be assisted by possible recombination events which refine derived haplotypes in the mixed-breed dogs (see Chapter 4).

In this chapter, high-resolution 3D reconstructions of the canine skull have been generated from computer tomography (CT) scans. Computational landmarking of the skull reconstructions has negated possible errors in the two-stage landmarking and merging of ventral and dorsal datasets which was required when using a MicroScribe (Schoenebeck *et al.* 2012). Furthermore, this chapter has included the use of mixed-breed dogs to exploit their diverse morphological variation. Finally, new methods to counteract the detrimental confounding effects of allometry have been developed to provide a best-practice approach to analysing canine morphologies.

3.2 Materials and Methods

3.2.1 Participants

The R(D)SVS Veterinary Ethical Review Committee (20 16, University of Edinburgh) approved the study which recruited participants from 2008 to the present from the Hospital for Small Animals, The University of Edinburgh. Archived records of patient computer tomography (CT) scans were searched using the ClearCanvas Workstation (v2.0.12729.37986) and matched with biomaterial archived and stored at -20 °C. If frozen biological samples were not available, a priority list of patients to contact was created. Potential study participants were targeted if they represented a morphological diverse breed, or were a breed at risk of developing brachycephalic obstructive airway syndrome (BOAS). A candidate list of dogs was provided to Professor Richard Mellanby and Susan Campbell who made contact with owners directly. If owners were willing to participate in the study, a PERFORMAgene™ saliva swab with instruction for its use, a consent form, a prepaid return envelope and a short summary of the project aims were sent to owners. Genomic DNA (gDNA) was extracted from the returned saliva swabs as previously described (see Section 2.3) and stored at 4 °C for later use in Chapter 4. From 2010 onwards, the owners of new patients receiving a diagnostic CT scan were asked during their consultation whether they would be willing to be included in the study. Additionally, participants were recruited from veterinary practices across the UK and Switzerland and included: Davies Veterinary Specialists, Hertfordshire, UK; Small Animal Medicine and Surgery Group, The Royal Veterinary College, Hertfordshire, UK; and The Division of Clinical Radiology, The Vetsuisse Faculty of Bern, Switzerland.

All canine participants were admitted to their respective veterinary practices and received computer tomography (CT) scans as part of their clinical diagnosis. Common diagnoses included foreign bodies, ear infections, BOAS and physical trauma to the skeleton. A Siemens Somatom four-slice CT scanner was used to acquire spiral or sequential scans of either one- or two-millimetre slice thickness. All scans were reviewed by a veterinary specialist in diagnostic imaging as part of their

diagnosis. Case notes were subsequently reviewed to confirm the suitability for the inclusion of cases in this study – dogs representing breeds with diverse craniofacial morphologies, patients diagnosed with BOAS and breeds with an increased susceptibility of BOAS were prioritised. The only exclusion criteria were cases with pathologies or injuries compromising the structure of the skull. Later, during isosurface reconstruction, several study candidates were excluded as a result of corrupt archived files which failed to generate an isosurface.

Due to the ongoing recruitment of patients and the continued archiving of biological samples, participants have been grouped into two study cohorts. The first study cohort was analysed and provided the resolution of the major locus controlling canine brachycephaly (discussed throughout this thesis) which was published in 2017 (Marchant *et al.* 2017). Subsequently, the first study cohort was extended to form the second study cohort which has provided more statistical power and allowed for the modification of study procedures. Throughout this thesis, the use of study cohorts one and two are stated clearly for each analysis.

3.2.1.1 Cohort 1 Demographics

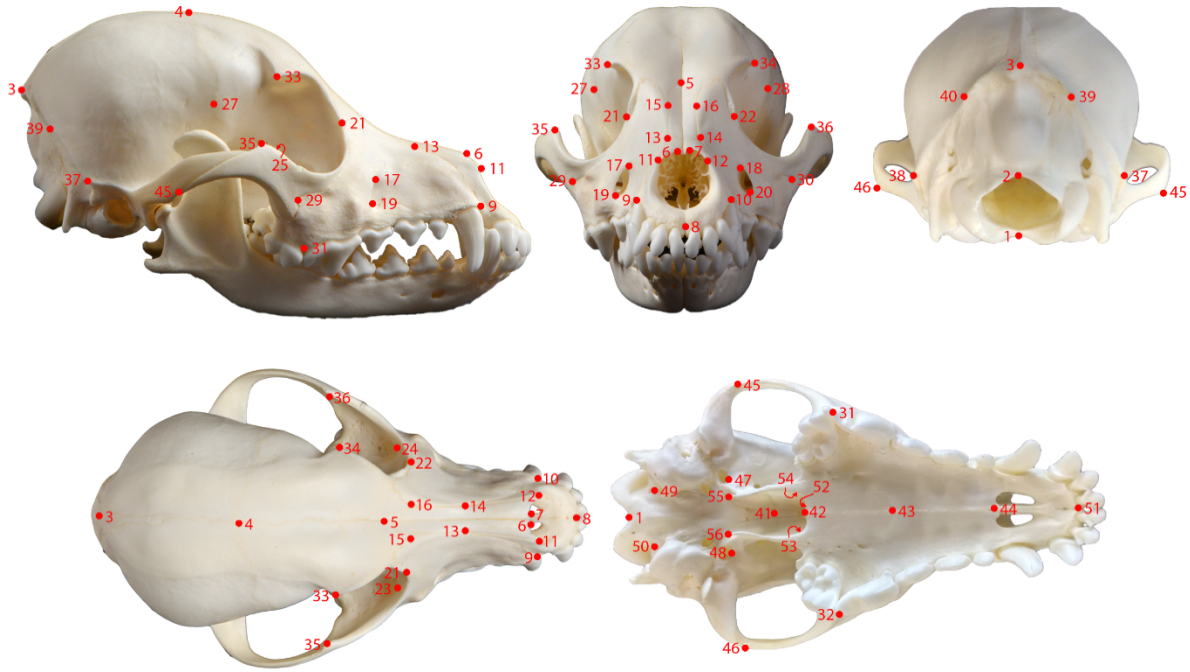
In total, the first cohort consisted of 374 (212 male, 162 female) referral patients each with computer tomography (CT) scans. All individuals were 24 months old and above to ensure the complete maturation of the bones of the skull. Of the 374 dogs, 291 represented 84 breeds recognised by the UK Kennel Club with a median of 2 dogs per breed and a minimum of 1 (various breeds) and a maximum of 16 (Labrador Retrievers & Golden Retrievers). In addition to this, a further eighty-three mixed-breed dogs were included. The neurocranium and viscerocranium was examined for all participants however a subset of 355 dogs were used for the analysis of the mandible. Nineteen dogs could not be used for the mandible assessment due to physical trauma to the mandible or absence of the mandible from the CT scan which were often focussed on just the cranium for diagnostic purposes.

3.2.1.2 Cohort 2 Demographics

All participants of the first study cohort were included in the second cohort. Additionally, a further 186 dogs were added to give a total of 560 dogs (315 males, 245 females). Breed diversity was increased to include a total of 92 UK Kennel Club recognised breeds with a median of 2 dogs per breed and a minimum and maximum of 1 (various breeds) and 38 (Labrador retriever), respectively. A total of 114 mixed-breed dogs were also included in the cohort. All dogs were aged 24 months and above. The neurocranium and viscerocranium was analysed for all participants however the mandible was no longer assessed. Throughout this thesis, the breed designations of participants were owner defined but were later verified by genotypic profiling (see Chapter 4).

3.2.2 Skull Reconstruction and Landmarking

High-resolution three-dimensional (3D) skull reconstructions of each patient were reconstructed using the CT scan-derived DICOM files in the Stratovan Checkpoint software (v2016.06.28.0428). Once an individual's DICOM stack had been loaded into the Checkpoint software, the "Surface" tab was selected. To display bone, a Hounsfield unit of approximately 500 was used. Anatomical substructures (cranium and mandible) of the resulting skull isosurfaces were landmarked by a single analyst. Patient demographics information was hidden from the analyst. Fifty-six landmarks were selected to capture the overall variation in the cranium whilst a further thirty landmarks were used for the mandible (Figure 3.1 and 3.2). Landmark positions were selected based upon their reproducibility (Schoenebeck *et al.* 2012). Once landmarking was complete, the Stratovan software generated a .ckpt save file which contained raw 3D coordinates for each landmark. A custom-made R script extracted the raw coordinates for all study participants and created a reformatted plain text (.txt) file containing x, y and z coordinates for every landmark across all participants. The Hounsfield units used to reconstruct skull isosurfaces were also extracted as metadata for later use.

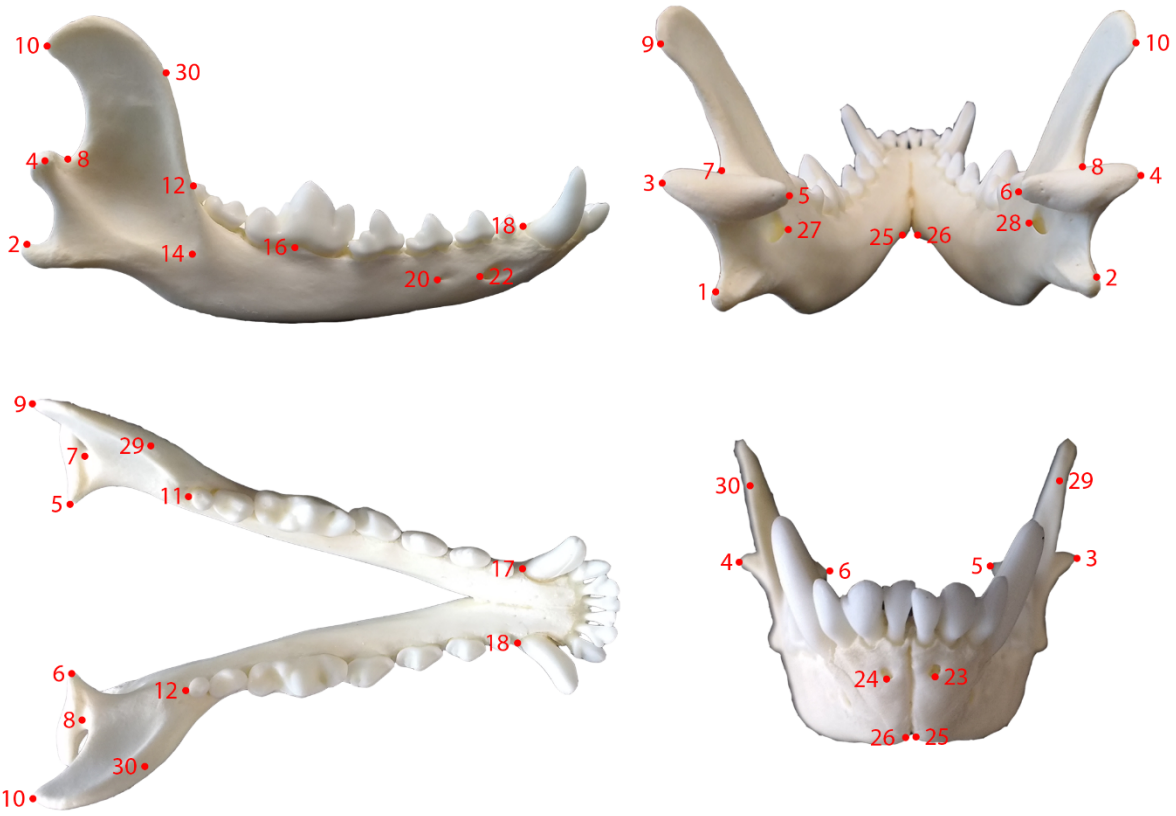


1. Basion (Posterior Hinstammbasis) *
2. Opisthion *
3. Inion *
4. Bregma *
5. Nasion *†
6. Internasal Suture (right) †
7. Internasal Suture (left) †
8. Prosthion †
9. Incisive-Maxilla Junction with Alveolus (right) †
10. Incisive-Maxilla Junction with Alveolus (left) †
11. Nasal Bone Process (right) †
12. Nasal Bone Process (left) †
13. Incisive-Maxilla-Nasal Junction (right) †
14. Incisive-Maxilla-Nasal Junction (left) †
15. Frontal-Maxilla-Nasal Junction (right) †
16. Frontal-Maxilla-Nasal Junction (left) †
17. Dorsal Infraorbital Foramen (right) †
18. Dorsal Infraorbital Foramen (left) †
19. Ventral Infraorbital Foramen (right) †
20. Ventral Infraorbital Foramen (left) †
21. Frontal-Lacrimal-Maxilla Junction (right) †
22. Frontal-Lacrimal-Maxilla Junction (left) †
23. Dorsal Maxillary Foramen (right)
24. Dorsal Maxillary Foramen (left)
25. Dorsal Ethmoid Foramina (right)
26. Dorsal Ethmoid Foramina (left)
27. Pterion (right) *
28. Pterion (left) *
29. Zygomatic Process of the Maxilla (right) †
30. Zygomatic Process of the Maxilla (left) †
31. Molar-Premolar Interalveolar Septa (right) †
32. Molar-Premolar Interalveolar Septa (left) †
33. Zygomatic Process of the Frontal Bone (right) *
34. Zygomatic Process of the Frontal Bone (left) *
35. Frontal Process of the Zygomatic Bone (right)
36. Frontal Process of the Zygomatic Bone (left)
37. Nuchal Crest Flare (right) *
38. Nuchal Crest Flare (left) *
39. Asterion (right) *
40. Asterion (left) *
41. Anterior Hirnstammbasis *
42. Posterior Nasal Spine
43. Maxilla-Palatine Junction at Midline †
44. Incisive-Maxilla Junction at Midline †
45. Ventral Temporal-Zygomatic Junction (right)
46. Ventral Temporal-Zygomatic Junction (left)
47. Rostral Foramen Ovale (right) *
48. Rostral Foramen Ovale (left) *
49. Rostral Hypoglossal Canal (right) *
50. Rostral Hypoglossal Canal (left) *
51. Intraoral Midline of Incisive Bone at Alveolus †
52. Midline of Vomer Dorsal to Nasal spine
53. Midline of Choana Dorsal to Posterior Nasal Spine (left)
54. Midline of Choana Dorsal to Posterior Nasal Spine (right)
55. Pterygoid Hamulus (left)
56. Pterygoid Hamulus (right)

Figure 3.1 - Landmarks of the Cranium. Position of landmarks placed on the reconstructed isosurfaces of the skull divided into landmarks of the viscerocranium (*) and neurocranium (†). Images clockwise from top-left: Lateral (right side), Rostral, Caudal and Ventral Dorsal. Modified from (Marchant *et al.* 2017).

3.2.3 Morphometric Analysis

The reformatted 3D coordinate file was imported into MorphoJ (v1.06c, Klingenberg Lab) (Klingenberg 2010). A generalised Procrustes fit was performed to appropriately superimpose the shape data by scaling, transposing and rotating



1. Dorsal Angular Process (left)
2. Dorsal Angular Process (right)
3. Lateral Condylar Process (left)
4. Lateral Condylar Process (right)
5. Medial Condylar Process (left)
6. Medial Condylar Process (right)
7. Dorsal Mandibular Notch (left)
8. Dorsal Mandibular Notch (right)
9. Dorso-Caudal Coronoid Process (left)
10. Dorso-Caudal Coronoid Process (right)
11. 3rd Molar Caudal-midline at Alveolus (left)
12. 3rd Molar Caudal-midline at Alveolus (right)
13. Ventral Masseteric Fossa (left)
14. Ventral Masseteric Fossa (right)
15. 1st Molar Lateral-midline at Alveolus (left)
16. 1st Molar Lateral-midline at Alveolus (right)
17. Canine Tooth Caudal-midline at Alveolus (left)
18. Canine Tooth Caudal-midline at Alveolus (right)
19. Caudal aspect of Caudal Mental Foramina (left)
20. Caudal aspect of Caudal Mental Foramina (right)
21. Caudal aspect of Middle Mental Foramina (left)
22. Caudal aspect of Middle Mental Foramina (right)
23. Caudal aspect of Rostral Mental Foramina (left)
24. Caudal aspect of Rostral Mental Foramina (right)
25. Ventral Aspect of Inter-Mandibular Joint (left)
26. Ventral Aspect of Inter-Mandibular Joint (right)
27. Rostral Mandibular Foramen (left)
28. Rostral Mandibular Foramen (right)
29. Mid-Coronoid Process (left)
30. Mid-Coronoid Process (right)

Figure 3.2 – Landmarks of the Mandible. Position of landmarks placed on the reconstructed isosurfaces of the skull. Images clockwise from top-left: Caudal, Lateral (right side), Dorsal and Rostral. Modified from (Marchant *et al*. 2017)

(Klingenberg 2016). Landmark pairs were checked for errors using the "Find Outliers" function. The landmarks of the cranium were subdivided into those representing the neurocranium (n = 18) and viscerocranium (n = 25) (Figure 3.1 and 3.2). The viscerocranium was defined by structures rostral to, and not inclusive of the frontal bones. The neurocranium was defined by all structures caudal to this. In order to remove the effects of allometry from the whole head, viscerocranium and mandible symmetric coordinate data, a regression of 10,000 permutations using the neurocranium centroid size was used. The neurocranium centroid is a bi-product of the Procrustes fit which indicates the scaling required to superimpose the neurocranium shape data and has been used as a proxy to body size previously (Schoenebeck *et al.* 2012).

Subsequently, in an attempt to best counteract the effects of allometry, two additional methods were trialled in the second study cohort. The first method used a pooled within-group regression for breed. Here, the study population was split into its constituent breeds and within these groups, shape (whole head and viscerocranium PC1) was regressed on neurocranium centroid size. The rationale for doing this is that size is expected to be structured by breed and so a single linear regression for the entire dataset may be less representative than a within breed regression (Klingenberg 2016).

A second method similarly used a within-group regression, however this time the groups were defined by the combination of size-associated QTL carried by the study participant. To identify these size-related QTL, the neurocranium centroid size from the second study cohort was used as a phenotypic input for a GWAS in Chapter 4. Nine QTL were associated and at positions of previously associated variants, individuals were genotyped for the variant (see Section 2.5.6), whilst in the absence of a variant, the index marker (most significant marker) for each QTL was used. The rationale behind this was to predict the size of an individual and negate possible errors in phenotype capture, environmental influences and assumptions of homogeneity across breeds (as used in the previous method).

Following each of the regression methods to remove allometry, a covariance matrix was generated from the regression residuals and a principal component analysis (PCA) was used to reduce dimensionality. The prebasial angle between the plane of the cranial base (hirnstambasis) and the plane of the hard palate was calculated using a custom R script. Briefly, the symmetric coordinates from the Procrustes fit in MorphoJ were used as an input. An assumption was made that there was negligible variation in the z-axis of coordinates since all landmarks are situated along the midline of the x/y-axis, therefore 2D vectors were used to calculate the prebasial angle. Angles are given in degrees.

3.3 Results

3.3.1 Skull Reconstructions

CT scans for each study participant were reconstructed to produce a high-resolution three-dimensional (3D) isosurface of the skull completely absent of soft

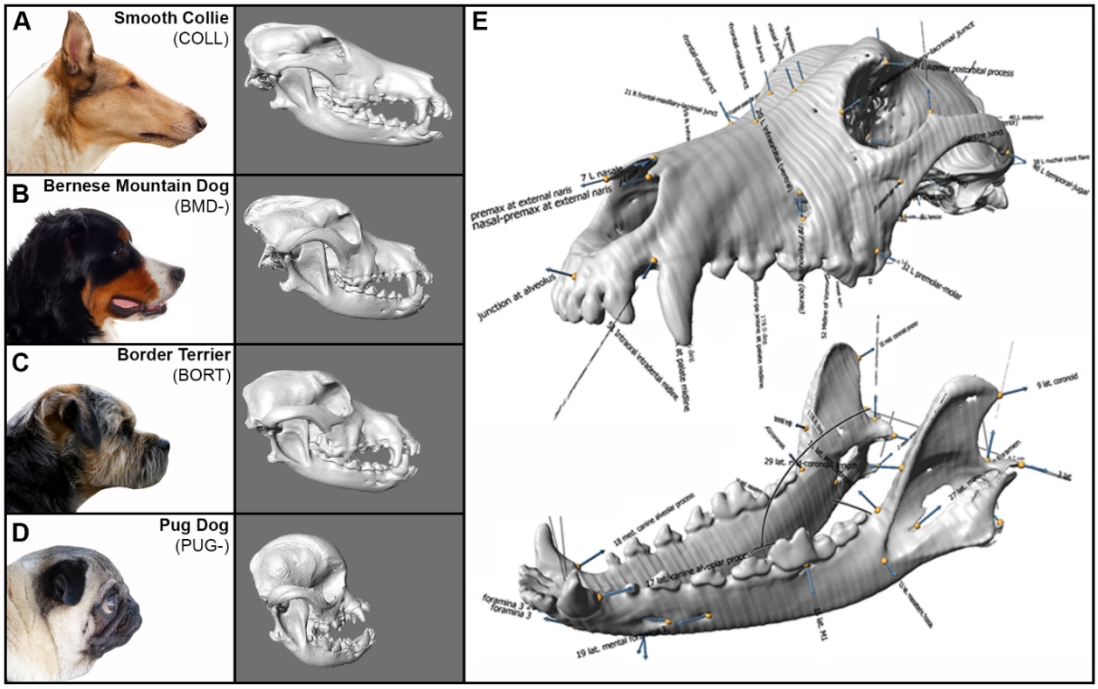


Figure 3.3 – Isosurfaces of the Canine Skull. Lateral images of a Smooth Collie (A; dolichocephalic), Bernese Mountain dog (B; mesocephalic), Border Terrier (C; mesocephalic) and a Pug (D; brachycephalic) with corresponding isosurfaces of the skull. Images and reconstructions are not to scale. (E) The cranium and mandible are separated to reveal the position of landmarks across the two anatomical structures. Modified from (Marchant *et al*. 2017).

tissue or trimmings which may hamper other phenotyping approaches (Figure 3.3). Extremes of craniofacial phenotypes can be seen in the reconstructions such as the brachycephalic skull shape of the Pug and dolichocephaly in the Smooth Collie. The Bernese mountain dog and Border terrier represent breeds with an intermediate skull shape. Eighty-six landmarks (56 cranial, 30 mandibular) were placed across the skull isosurfaces by a single analyst to capture the subtle and extreme morphological variation seen both within and across breeds (Figure 3.1, 3.2 & 3.3).

3.3.2 Geometric Morphometrics on Cohort 1

To remove the potential confounding effects of differences in orientation, position and scale of the skull coordinates, a Procrustes fit for the whole skull was executed. Subsequently, a regression using the neurocranium centroid size on the symmetric Procrustes coordinates then removed the influence of size on form (allometry) to leave variation in shape alone. Finally, a principal component analysis was performed on the regression residuals.

Whole head PC1 explains up to 63.52% of the total variation in the skull across study cohort 1 with each subsequent PC explaining an ever-reducing percentage of the variation (Table 3.1). Global changes in the morphology of the whole head are visualised using wireframe schematics (Figure 3.4A). As whole head PC1 becomes more negative the height of the cranium increases, the width of the skull broadens and the length of the skull becomes shorter in the rostrocaudal axis – all

Principal Component	% Variance Explained		
	Whole Head	Viscerocranium	Mandible
1	63.52	72.24	66.20
2	8.82	5.15	8.54
3	4.95	4.07	3.02
4	2.29	2.41	2.60
5	1.66	2.18	2.31

Table 3.1 – Principal Component Weightings for Cohort 1. Principal components successively explain smaller tranches of morphological variation (percentage) within the distance matrix generated from each of the whole head, rostrum and mandible.

characteristics of the brachycephalic head conformation. Conversely, as whole head PC1 becomes more positive, the height of the cranium is reduced, the width of the skull becomes narrower and the length of the skull becomes longer – all characteristics of the dolichocephalic head conformation. A lollipop schematic for the whole head PC1 reveals the extent and direction of variation of each individual landmark from its mean position across the dataset (Figure 3.4B). Landmarks constituting the viscerocranium display the greatest variation relative to the rest of

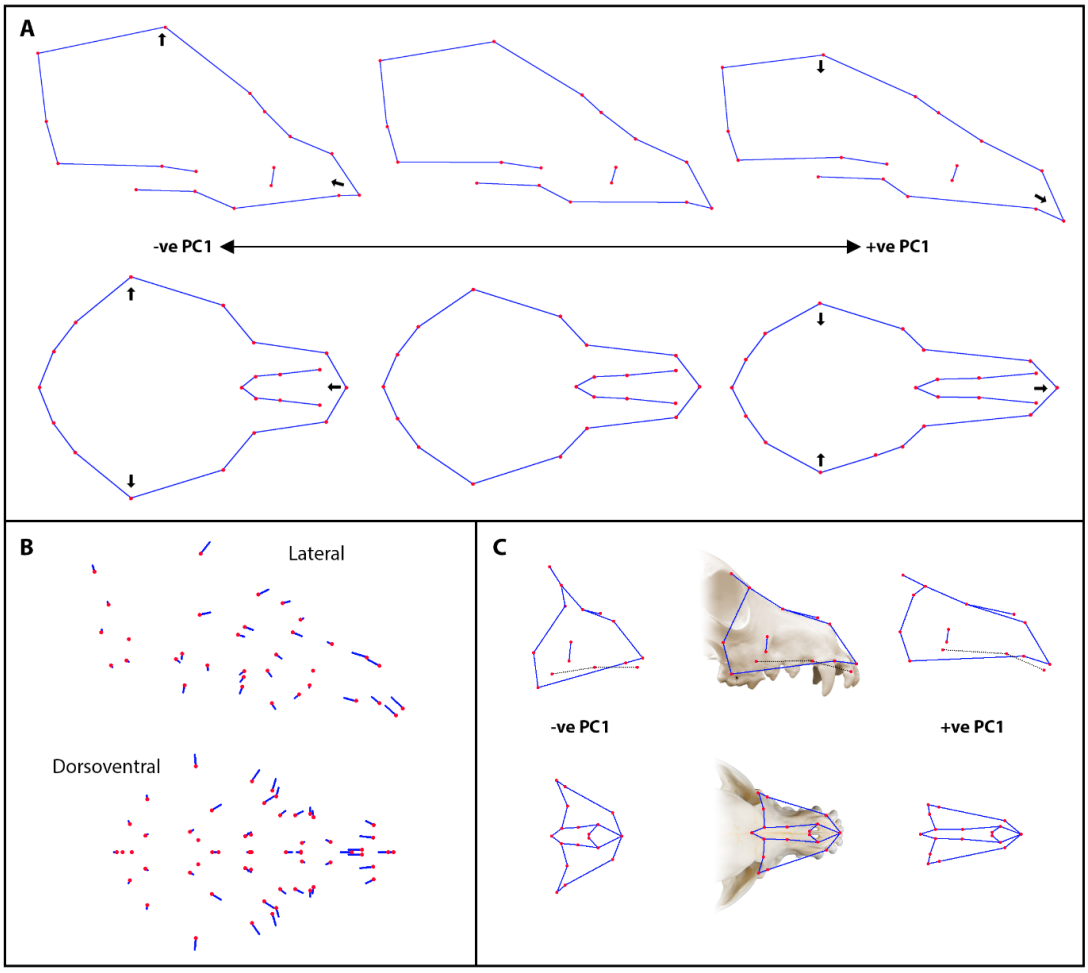


Figure 3.4 – Geometric Morphological Variation. Lateral and dorsoventral schematics of the (A & B) whole head and (C) viscerocranium summarising the geometric changes captured by principal component analysis. Wireframe diagrams depict the global geometric changes captured by negative and positive principal component 1 ("ve PC") for the (A) whole head and (C) viscerocranium. Red circles indicate surface landmarks of the skull. Connecting blue lines are added to provide visual context to shape changes. (B) Lollipop diagram shows the mean position of each landmark (red circles) and the direction and extent of variation captured across the cohort by PC1 (blue lines).

the skull. To this end, the geometric analyses were repeated using the viscerocranium as an anatomical subset. Viscerocranium PC1 displayed the similar global morphological variation with negative PC1 accounting for a broader, shorter face whilst positive PC1 accounted for a longer, narrower face (Figure 3.4C). These are characteristics of brachycephaly and dolichocephaly, respectively. Despite capturing similar morphological trends, viscerocranium PC1 accounts for a greater percentage of the total variation across the foreface (72.24%) than whole head PC1 accounted for across the skull (63.52%) (Table 3.1). This suggests a modular localisation of variation such that the greatest morphological variation is restricted to the viscerocranium compared to the skull as a whole. The remaining principal components (PC2-5) are difficult to interpret due to the reduced level of phenotypic variation explained (data not shown). Later, the larger study cohort 2 will be used to interpret morphological changes captured by whole head PC2.

The neurocranium centroid size was generated as a bi-product of the Procrustes fit and used in subsequent analyses as a proxy for body size. Here, the neurocranium centroid size appears to have minimal variation within breeds and gives an expected order to breeds within the cohort i.e. the miniature pinscher, toy poodle and Japanese Chin are some of the smallest breeds in this study which is reflected in their small neurocranium centroid size (Figure 3.5A). Conversely, the giant breeds which include breeds such as the Newfoundland, Great Dane and Leonberger have some of the largest neurocranium centroid sizes and appear to have an inflection in their neurocranium centroid sizes compared to the rest of the breed distributions. The distributions of PC1 for the whole head, viscerocranium and mandible all show a similar trend – brachycephalic breeds have negative PC1 scores and have a distinct inflection in the breed boxplot distributions compared to other breeds (Figure 3.5B-D). This reflected in the non-normal, bi-modal frequency distributions for each trait (Figure 3.5B-D insert).

The neurocranium centroid size and PC1 scores for each anatomical substructure (whole head, viscerocranium and mandible) described here (Figure 3.5)

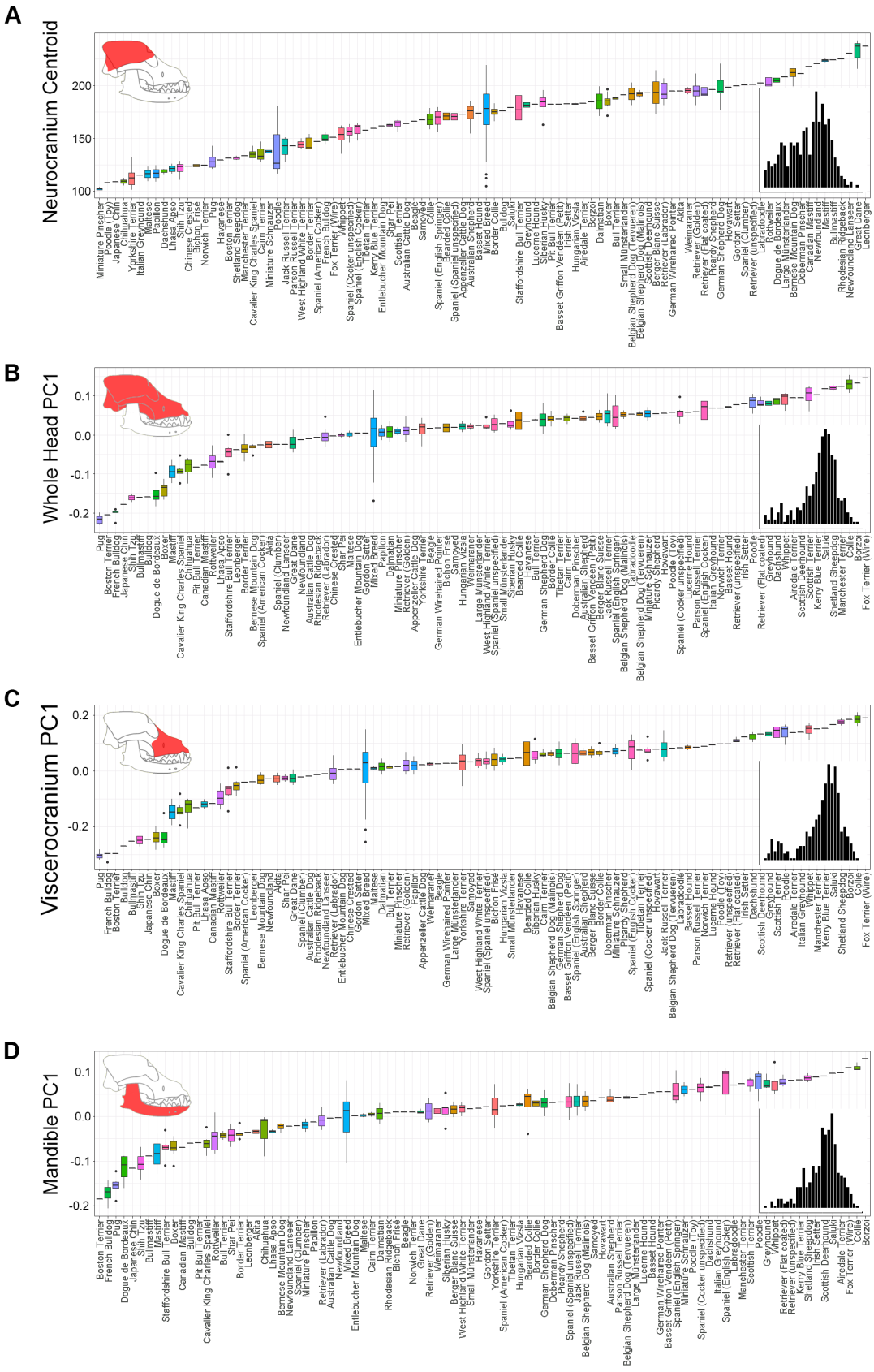


Figure 3.5 – Morphology by Anatomical Subdivision. Boxplots of the (A) neurocranium centroid, (B) whole head PC1, (C) viscerocranium PC1 and (D) mandible PC1 for each breed. Insert: frequency distributions.

are used as phenotypic inputs for the genome-wide association analyses in Chapter 4.

3.3.3 Geometric Morphometrics on Cohort 2

3.3.3.1 Methods to Remove Allometry

Following the increased availability of study participants in the extended study cohort 2, new methods to account for allometry in the dataset were sought to increase the power of subsequent analyses. The neurocranium centroid size has proven to be an important proxy for body size and began to address removing the influence of allometry on shape variation in the skull. Despite this, the regression scores of the viscerocranium remain in strong correlation with neurocranium centroid size (R^2=0.889), suggesting a component of size still remains in the dataset (Figure 3.6A). Here, two additional approaches to account for allometry were tested which used regressions pooled by subgroup including breed (Figure 3.6B) and combination of size-associated genetic markers (Figure 3.6C). Prior to this regression, the neurocranium centroid from the second cohort was applied in a genome-wide association analysis (Chapter 4). This resulted in nine QTL being associated with body size. The index marker of each of these loci were used as size-associated markers as

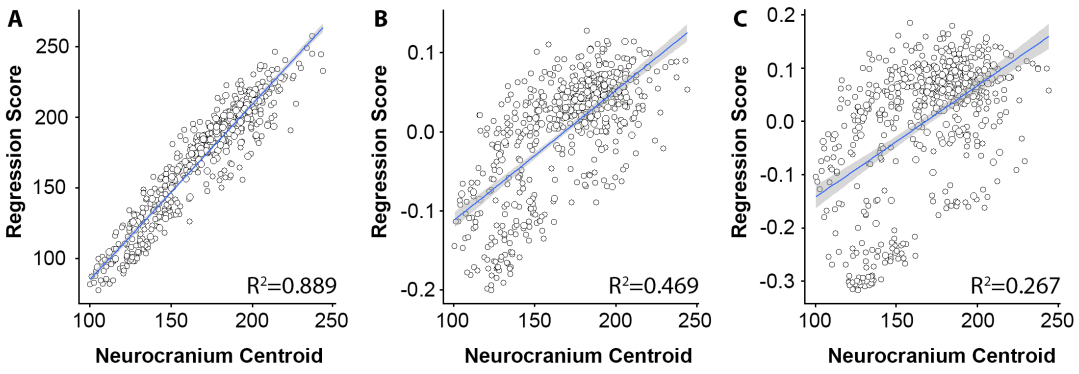


Figure 3.6 - Parsing Shape from Size. The (A) neurocranium centroid size, (B) neurocranium centroid size pooled by breed and (C) the neurocranium centroid size pooled by nine morphological size-related genotypes were used to regress the aspect of size from the viscerocranium. The resulting regression scores were plotted against the neurocranium centroid size. Pearson's correlation coefficient is given.

part of the pooled regression. The extent of correlation between the neurocranium centroid size and the regression scores is reduced for both the breed (0.469) and size-associated marker (0.267) indicating a reduced influence of allometry on the dataset. For this reason, the pooling of individuals by size-related markers as a method to account for allometry was used in subsequent studies of the whole head and viscerocranium.

3.3.3.2 Principal Component Analysis

Following the pooled regression described above, a principal component analysis was performed for both the whole head and viscerocranium datasets in study cohort 2. For the second cohort, there was an increase in the percentage of phenotypic variation explained by PC1 in both the whole head and viscerocranium dataset when using the pooled regression instead of universal regression (Table 3.2). The type of regression used seemed to have a minimal impact on the remaining principal components.

Principal Component	Single Neurocranium Centroid Regression (old method)	Within-group Pooled Regression by Morphological QTL (new method)
Whole Head 1	70.282	71.701
Whole Head 2	5.12	4.82
Whole Head 3	2.661	2.56
Whole Head 4	2.369	2.327
Whole Head 5	1.799	1.717
Viscerocranium 1	78.823	79.962
Viscerocranium 2	3.553	3.359
Viscerocranium 3	3.306	3.075
Viscerocranium 4	1.732	1.644
Viscerocranium 5	1.561	1.453

Table 3.2 – Principal Component Weightings for Cohort 2. Principal components successively explain smaller tranches of morphological variation within the distance matrix generated from each of the whole head and viscerocranium. Regressions have been performed using a single neurocranium centroid across the entire dataset (old method) and a within-group regression whereby individuals are pooled by their combination of size-related QTL (new method).

Plotting viscerocranium PC1 scores against the neurocranium centroid size allows for the morphological segregation of breeds. Breeds such as the French Bulldog and Pug cluster tightly together whilst breeds such as the Yorkshire Terrier and Labrador Retriever appear to have a greater degree of craniofacial variability (Figure 3.7A). Importantly, this breed-level clustering can allow for the visual inspection of the data to confirm breed designation. For example, a dog was labelled as a Smooth Collie however its viscerocranium PC1 score appeared as an outlier and so prompted the checking of its clinical records which confirmed that the dog was in fact a mixed-Collie dog (black arrows). The frequency distribution of viscerocranium PC1 scores forms a distinct bi-modal distribution with the "classic" brachycephalic

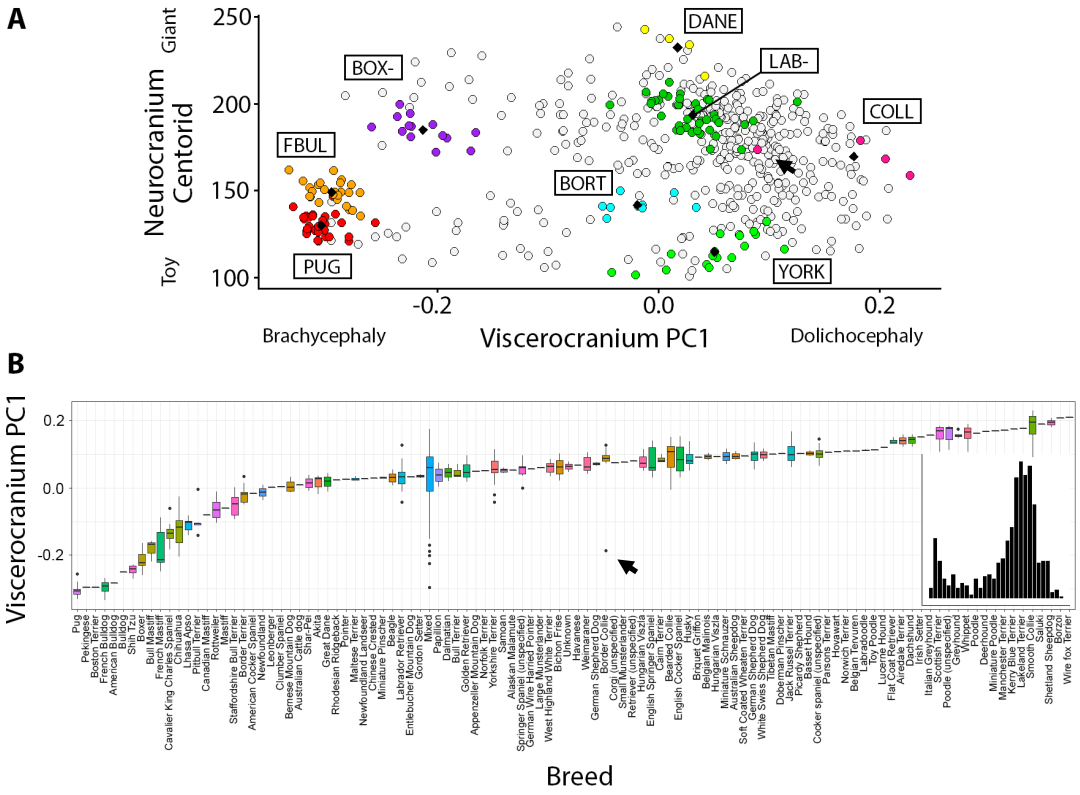


Figure 3.7 – Second Study Cohort Phenotype Distribution. (A) Distribution of viscerocranium PC1 values and neurocranium centroid scores for the second phase cohort. Example breeds are highlighted. Breed abbreviations: Pug (PUG), French Bulldog (FBUL), Boxer (BOX-), Great Dane (DANE), Labrador Retriever (LAB-), Smooth Collie (COLL), Yorkshire Terrier (YORK) and Border Terrier (BORT). (B) Breeds ordered by viscerocranium PC1 score. Black arrow indicates a dog incorrectly recorded as a Border Collie which was later found to be a mixed-Collie dog.

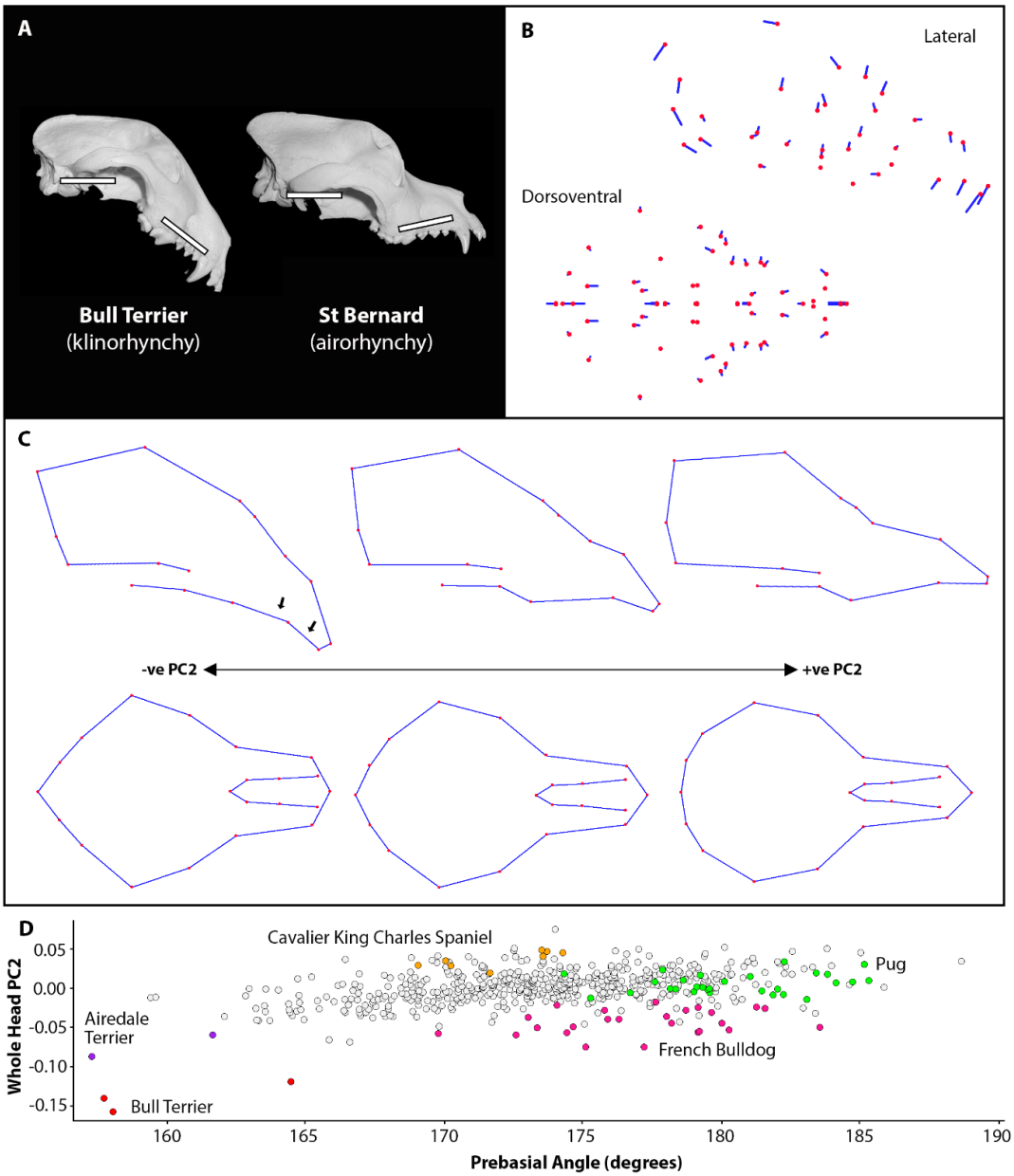


Figure 3.8 – Whole Head PC2. (A) Lateral photographs of the Pitbull Terrier and St. Bernard dog skulls which represent klinorhynchy and airorhynchy, respectively. Rectangular boxes indicate the planes of the cranial base (left) and hard palate (right). (B) Lollipop diagrams of lateral and dorsoventral profiles of the whole head show the mean position of each landmark (red circles) and the direction and extent of variation captured across the second cohort by whole head PC2 (blue lines). (C) Wireframe diagrams depict the global changes in the lateral (top) and dorsoventral (bottom) outline of the whole head capture by positive and negative PC2. (D) Distribution of individual dogs scored by whole head PC2 and prebasial angle calculated from the geometric landmark data. Example breeds are coloured and appear to indicate a greater variability in prebasial angle than whole head PC2.

breeds contributing the majority of the distinct population (Figure 3.7B). This is phenomenon is reflected in the boxplot distribution which sees a distinct inflection in the distribution for the brachycephalic breeds.

The increase in population size for study cohort 2 and the inclusion of more diverse breeds has allowed for the interpretation of additional PCs such as whole head PC2. Visual inspection of whole head PC2 indicates it is likely capturing the variability in the angle of the hard palate relative to the neurocranium. The whole head PC2 axis is polarised by the Bull Terrier and Airedale Terrier which are breeds displaying klinorhynchy whilst breeds such as the Cavalier King Charles Spaniel have a greater extent of airorhynchy as has been shown in previous studies (Mitchell *et al.* 2014; Knowler *et al.* 2017). The extremes of this klinorhynchy-airorhynchy axis can be seen in breeds such as the Bull Terrier and St Bernard, respectively (Figure 3.8A). Lateral lollipop plots of the whole head indicate that landmark variation occurs in the foreface relative to the neurocranium (Figure 3.8B). The effects of this are more easily visualised in the wireframe diagrams (Figure 3.8C). Here, the lateral outline of the skull can be seen to have a severely "downward-tilted" foreface as whole head PC2 becomes more negative (klinorhynchy) whilst it is dorsally rotated as whole head PC2 becomes more positive (airorhynchy). Meanwhile, the shape of the neurocranium appears to have minimal variation as the PC2 value varies. Previous studies have relied on the use of the prebasial angle to determine the extent of klinorhynchy/airorhynchy however here, the prebasial angle (degrees) has a greater variation within breeds than whole head PC2 (Figure 3.8D). This could suggest that the prebasial angle is a poor representation of the morphological variation in the palate angle and that a global analysis of the foreface relative to the neurocranium is required.

3.3.3.3 The Use of Mixed-Breed Dogs

For the first time, mixed-breed dogs have been used to assess canine craniofacial morphological variation. All morphological traits shown here (neurocranium centroid size, viscerocranium PC1, whole head PC2 and prebasial

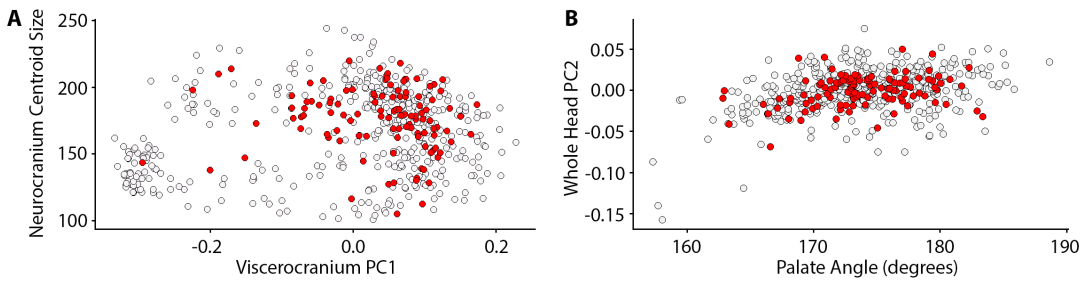


Figure 3.9 – Mixed-breed Dog Population. Distribution plots of (A) neurocranium centroid size against viscerocranium PC1 and (B) whole head PC2 against the prebasial angle for the second study cohort. Mixed-breed dogs are coloured in red and occupy a smaller morphospace which is entirely encapsulated by that of their purebred counterparts.

angle) indicate that all mixed-breed dogs (red circles) occupy a smaller morphospace which exists entirely within that of the purebred population (Figure 3.9). At this morphological level, the mixed-breed population display similar characteristics to their purebred counterparts however it is anticipated that this will be the result of unique combinations of underlying genetic loci which will provide more power when later modelling the phenotypic effect size of these loci (see Chapter 4). Hopefully, the mixed-breed population will have also undergone more extensive recombination events than the purebred population which will allow for the refinement of derived haplotypes in later analyses.

3.3.3.4 Sexual Dimorphism

Sexual dimorphism is a common phenomenon across biology. To determine whether this is true across dogs in this dataset, the neurocranium centroid size was plotted by sex for the five most populated breeds – French Bulldog (25), Golden Retriever (26), Jack Russel terrier (18), Labrador Retriever (38) and Pug (29) (Figure 3.10). All five breeds displayed significant sexual dimorphism with the male being larger than the female in each instance. Neuter data for all dogs was not available. Sex can be incorporated as a covariate for downstream genomic analysis.

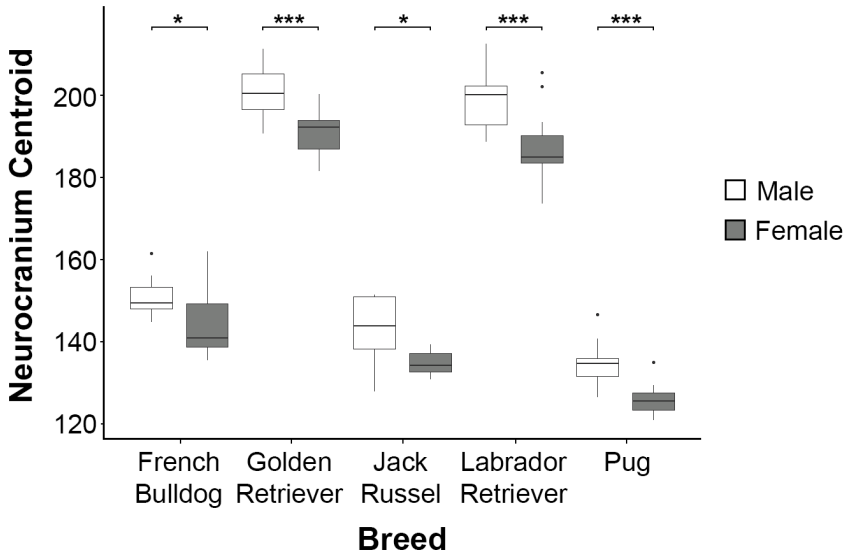


Figure 3.10 - Sexual Dimorphism. Male (white) and female (grey) neurocranium centroid sizes for example breeds displaying intrabreed sexual dimorphism (t-test: <0.05 *; <0.001 *).**

3.4 Discussion

The results in this chapter support previous observations that the domestic dog is a highly morphologically diverse species. The establishment of the breed clubs within the last 200 years followed by heavy artificial selection has resulted in the exaggeration of morphological traits, which in many cases, have become breed defining (Wayne and Ostrander 2007; Club 2018b). Here, the focus was specifically on the skull and the variation in both its shape and size. New methods were sought to define and score this morphological variation with the ultimate goal of using these morphological scores to later identify their underlying genetic contributors in subsequent genome-wide association analyses.

The data in this chapter have highlighted 3 major axes of morphological variation in the canine skull: brachycephaly-to-dolichocephaly, klinorhynchy-to-airorhynchy and small-to-large morphological variation. Beyond this, visual inspection of the skulls suggest that variation also exists in other craniofacial features such as the curvature of the neurocranium, the formation of a neural ridge, the arches of the zygomatic bones, the angle of the mandibular arm (ramus), the depth

of orbits and several others which are beyond the scope of this study. Landmarks were less ideally positioned to characterise these traits which are generally more subtle than the 3 major axes listed above and so would warrant a new selection of landmarks to study in the future. An additional assessment of the current data could also be to reduce the total number of landmarks used in this study to capture skull variation as rather paradoxically, having a very large number of landmarks can in fact reduce the statistical power or ability to detect subtle shape differences (Gunz et al. 2013; Collyer et al. 2015).

The most prominent overarching axis of morphological variation described here has been the brachycephalic-to-dolichocephalic continuum in which skulls generally span from broad and short to long and narrow, respectively. These extremes of facial conformation were first described in human patients with craniosynostoses (Persing *et al.* 1989) but have since been observed across a wide range of mammalian species including across the wider Carnivora order (Sears *et al.* 2007), and in domestic species such as the cat (Künzel *et al.* 2003; Schlueter *et al.* 2009; Lyons *et al.* 2015), cow (Veitschegger *et al.* 2018), goat (Veitschegger *et al.* 2018), sheep (Dantas *et al.* 2013), pig (Veitschegger *et al.* 2018) and mice (Wang *et al.* 2005). This observation raises an interesting question whether or not a common pathway has been altered across species? The exploration of the genetic determinants of canine brachycephaly in Chapters 4 and 5 will begin to shed light on this.

Here, the geometric morphometric analysis of the whole canine skull revealed a pronounced variation in the viscerocranium. A subsequent PC analysis showed an increase in the percentage of total phenotypic variation explained by PC1 for the viscerocranium compared to the whole head suggesting some of the greatest variation in brachycephalic dogs is restricted to the foreface. Similarly, the whole head PC2 captured the klinorhynchy-to-airorhynchy axis in which the morphological variation is restricted to the viscerocranium relative to the neurocranium. This highlights a potential modularity in the control and development of the skull such

that the viscerocranium and neurocranium are correlated, yet distinct morphological substructures. The modularity of the skull is not a novel concept – Fondon and Garner, 2004 investigated the angle of the hard palate compared to the cranial base in the Bull Terrier and established that despite variation in the morphology of the viscerocranium, the neurocranium remained relatively unchanged (Fondon and Garner 2004). A study of St. Bernard's revealed that across all of the craniofacial variation, the majority was explained by an upward inclination of the palate relative the cranial base (airorhynchy) coupled with a broadening of the head, most notably around the nose (Drake and Klingenberg 2007). A subsequent study across almost 700 domestic dog skulls confirmed the morphological distinction in variation of the neurocranium and viscerocranium (Drake and Klingenberg 2010). The distinction of these anatomical substructures is perhaps unsurprising based on the knowledge of embryonic craniofacial development. The bone of the viscerocranium and neurocranium are largely sourced from two distinct populations in early embryogenesis – the neural crest and paraxial mesoderm, respectively (see Section 1.2). Differences in tissue origins could begin to explain the apparent disconnection in the behaviour of morphological variation across the two structures. It is possible to hypothesise that a genetic variant could specifically perturb a component of the viscerocranial developmental pathway whilst leaving the remaining neurocranial development relatively unaltered. Also, a possible explanation for the greater extent of viscerocranium variability compared to the neurocranium may be in part due to the tolerance of viscerocranial variation. Changes in the neurocranium, which houses the brain would be predicted to have more drastic effects on viability of offspring and so are less likely to be present in the variability of UK dogs.

The diverse variation in the overall size of the canine skull introduces the confounding influences of allometry when attempting to investigate the subtleties of craniofacial shape change. Schoenebeck et al., 2012 demonstrated that a failure to remove the allometric component in skull form variation results in the contamination of genomic association studies with size-related QTL, a phenomenon replicated in this study (data not shown) (Schoenebeck *et al.* 2012). Previous studies have used

breed average weight (kg) as an approach to remove allometry however this has the potential to fluctuate due to environmental factors and so may not truly reflect the influence of growth variation in the dog (Boyko *et al.* 2010). Superseding this was the use of centroid size regressions of the skull (Drake 2011; Schoenebeck *et al.* 2012). This use of a single linear regression was repeated here, however, the continued correlation between the neurocranium centroid size and the regression residuals of the viscerocranium suggested that an allometric component was still present in the data. Modifications to this procedure revealed that the pooling of individuals by their combination of size-related genetic markers improved the disentanglement of variation in craniofacial shape and its overall size (Klingenberg 2016). This procedure used the index markers from a genome-wide association test in Chapter 4 which were assumed to be tagging underlying causal variant(s). The advantages of this method is that it relieves the assumption that all dogs of the same size have the same combination of size-determining QTL and is thought to more accurately represent the changes in growth of individual dogs. A drawback of this approach is the requirement of a DNA sample for each dog – whilst this is not a concern for this study, it may prove to be a limitation for others. An interesting future experiment would be to repeat this protocol however, using definitive causal variants (once they are determined) rather than a tagging index marker which may not precisely capture the underlying variant genotype.

Overall, the data in this chapter importantly demonstrates that craniofacial morphological variation exists at both an inter-breed and intra-breed level. Purebred dogs have largely been assumed to be morphological homogeneous, however this data challenges that. It remains to be seen whether this observation is representative of the wider breed populations or if it is restricted to the participants of this study. An explanation for this variability could be that many of the dogs attending clinical practices across the UK represent 'family pets' which are not thorough-bred show-standard dogs and therefore liable to more extensive morphological variation. Another source of apparent intra-breed variability could be due to the owner-defined breed designations. The veterinary practices participating in this study record dog

breeds from their owners and so it is possible that, for example, an owner miss-recorded their dog as a Labrador Retriever when in fact it was a Golden Retriever. This mislabelling could begin to account for some of the apparent morphological variation within a designated breed however the likelihood of this is minimal due to the assessment of the genomic structure of all participants in Chapter 4.

The analyses in this chapter have formulated a new protocol to improve the accountability for the detrimental effects of allometry on craniofacial form. In light of this, craniofacial shape variation has been observed to vary between breeds as expected but also within breeds which challenges the pre-held assumption that a breed represents a morphologically homogeneous population. Importantly, this individual-level morphological quantification can permit the determination of the genomic contribution to these traits in Chapter 4.

Chapter 4: Genomic Assessment of Canine Morphology

4.1 Introduction

As explored in Chapter 3, the domestic dog represents a morphologically diverse species. This diversity, coupled with the simplified genomic architecture of the canine genome, have been exploited by many previous studies striving to find the genetic determinants of morphological variation.

4.1.1 Leg Morphology

Studies have explored a wide range of canine morphologies and shared success in identifying quantitative trait loci (QTL) or better still, putatively causal variants. One particular area of success has been in the mapping of leg length in dogs, namely in the form of the skeletal dysplasias. These are a broad collection of disorders that result in abnormal limb shortening such as that in asymmetric chondrodysplasia, a form of disproportional short stature that affects the long bones of the appendicular skeleton (Krakow 2015). A cross-breed analysis using a binary, long- versus short-legged phenotype discovered the insertion of a fibroblast growth factor 4 (*FGF4*) retrogene into canine chromosome (CFA) 18 (Parker *et al.* 2009). This retrogene was discovered to be expressed in a variety of breeds including the Miniature dachshund, Basset hound and Pembrokeshire Welsh corgi. Intriguingly, an independent study similarly used a short- versus long-legged phenotype for skeletal dysplasia to distil a second *FGF4* retrogene on CFA12 (Brown *et al.* 2017). This discovery was made using the Nova Scotia Duck Tolling Retriever but was subsequently shown to complement the breeds with skeletal dysplasia that did not have the CFA18 *FGF4* retrogene such as the American Cocker Spaniel, Beagle and French Bulldog. Owing in part to the simple Mendelian inheritance patterns in specific breed pedigrees, mutations in the collagen alpha-2(XI) chain (*COL11A2*), integrin subunit alpha 10 (*ITGA10*) and solute carrier family 13 (*SLC13A1*) have been found to cause forms of dwarfism in the Labrador Retriever, Norwegian Elkhound and Miniature Poodle respectively (Neff *et al.* 2012; Frischknecht *et al.* 2013; Kyöstilä *et al.* 2013). Furthermore, oculo-skeletal

dysplasia, a combination of dwarfism and eye defects is known to be caused by a mutation in the collagen alpha-3(IX) chain (*COL9A3*) in Labrador Retrievers (Goldstein *et al.* 2010).

4.1.2 Body Size Variation

This current study has a vested interest in the variation in skeletal size of dogs. Here, the neurocranium centroid size has been used as a proxy of skeletal size to remove the allometric component of craniofacial variation in the dog. An extensive literature exists for the hereditary component to body size variation in the dog, the studies on which have adopted a number of different approaches to investigating. One of the earliest and largest effect loci was identified by Chase et al., 2003. There, investigations were focused on a single breed – the Portuguese Waterdog, a breed permitted to vary up to 50% in size by the American Kennel Club (Chase *et al.* 2002). Principal components of measurements derived from two-dimensional radiographs of individual dogs were used as a phenotypic input (Chase *et al.* 2002). This study identified a quantitative trait locus (QTL) on CFA15 to be associated with small size. Later, a localised effort to resolve the CFA15 locus was able to define a critical interval encompassing the insulin-like growth factor 1 (*IGF1*) gene (Sutter *et al.* 2007). A definitively causal variant at this locus was unable to be isolated however within the critical interval is a short interspersed nuclear element (SINE) which is an appealing candidate for causality but it exists amongst numerous other potentially causal variants. Interestingly, the insulin-like growth factor 1 receptor (*IGF1R*) has since been associated with the “toy” breeds such as the Chihuahua and Miniature Poodle (Hoopes *et al.* 2012). A missense variant within this gene was discovered by parsing breeds based on averages of estimated heights into “toy” and “non-toy” categories. Care was taken to remove dogs with the CFA18 *FGF4* retrogene which induced asymmetric chondrodysplasia.

In a separate study, associated variants in a further four positional gene candidates were reported to correlate with standard breed weights averaged across male and female dogs (Rimbault *et al.* 2013). These candidates included the growth

hormone receptor (*GHR*) and stanniocalcin 2 (*STC2*) genes on CFA4 and the SMAD family member 2 (*SMAD2*) and high mobility AT-hook 2 (*HMGA2*) genes on CFA7 and 10, respectively. Allele frequencies at these four loci in addition to the *IGF1* and *IGF1R* were modelled to account for up to 52.5% of the total variance in canine body size suggesting many more body size QTL are yet to be identified. The extensive genetic contribution to canine body size variation was alluded to in studies by the Boyko Laboratory which used over 1,800 dogs to identify seventeen loci significant or suggestive associations with breed average male weight (Hayward *et al.* 2016). This list included the six aforementioned positional candidate genes and a collection of other appealing positional candidate genes and novel loci which together, are predicted to account for up to 80% of the total variance in canine body size. Fine mapping at these seventeen genetic loci remains to be done however, the extensive hereditary component to canine body size variation is clear.

4.1.3 Skull Shape Variation

Like body size, the complex nature of craniofacial shape in the domestic dog has been alluded to by several previous studies. One of the first studies to do this was by Jones et al., 2008 which took linear measurements of the skull and snout from dog photographs to generate 145 breed average ratios (Jones *et al.* 2008). This identified an association between snout length and a QTL on CFA1 at 97,045,173 (CanFam2). Later, Bannasch et al., 2010 also identified a CFA1 locus (~59.435 – 59.485 Mb) which has been the only analysis to date to examine haplotypes within the region (Bannasch *et al.* 2010). The study screened 49 SNPs across 88 dogs seemingly defined as brachycephalic which were compared to a selection of 185 non-brachycephalic breeds. Using a number of assumptions, a critical interval was drawn at the CFA1 locus which encompassed the thrombospondin 2 (*THBS2*) gene. Importantly, these assumptions may have been misguided. Firstly, their study expected that *all* brachycephalic dogs were fixed for the associated haplotype and therefore causal variant(s). Whilst this assumption is not unreasonable, the complex nature of the trait and known variability within breeds (see Chapter 3) may provide exceptions. Secondly, in reviewing the published haplotype analysis, it appears that the critical

interval may have been misinterpreted as the same haplotype appears to extend in both directions from the defined critical interval. No variant calling was performed within this interval which would be important if accurately genotyping at the locus is desired.

Emphasising the importance of this CFA1 locus with craniofacial variation in the dog, the locus was detected by Boyko et al., 2010 who used breed average allometric measurements of the skull. Association analyses revealed markers to surpass genome-wide significance across eight chromosomes however the allometric nature of the phenotype undoubtedly also returned cryptic associations with body size variation. Despite this, two clear associations were made which include the aforementioned CFA1 locus and a new locus at chr5:35,359,028 (CanFam2). Beyond this, no fine mapping or positional candidate genes were identified.

Finally, the most recent study assessing craniofacial variation in dogs was performed by Schoenebeck et al., 2012. This study is the only one to date to identify a putatively causal variant for canine brachycephaly. Specifically, a missense mutation in the bone morphogenetic protein 3 (*BMP3*) on CFA32 was identified in small brachycephalic breeds such as the Pekingese and Brussels griffon (Schoenebeck *et al.* 2012). This study used geometric morphometrics to analyse museum specimens to generate breed averages of skull shapes without allometry. Genotypes and phenotypes were disconnected as extant purebred show dogs were sampled as a source of DNA. Nonetheless, the power of this approach identified the only known putatively causal variant for canine brachycephaly.

4.1.4 Resolving the CFA1 Canine Brachycephaly Locus

The objective to this study was to address the shortcomings of the aforementioned works and in doing so, develop a new approach to isolating QTL and fine mapping them for complex canine morphologies. Specifically with regards to canine brachycephaly, collective drawbacks across these studies are observed in phenotyping, genotyping and the interpretation of the results.

Firstly, at a phenotypic level, several improvements to experimental design were implemented, as described in Chapter 3. In short, this included moving away from discrete phenotyping for head type. The use of "long" versus "short" face categorisation forces the assignment of dogs into two discrete groups which compromise the detection and resolution of the underlying QTL. Furthermore, the use of breed average phenotypes has its advantages, particularly when individual measurements are missing or in the presence of large environmental influences (e.g. body weight). However, in the instance of craniofacial shape variation, which has a higher hereditary component (Hunter *et al.* 1970; King *et al.* 1993; Johannsdottir *et al.* 2005), the use of breed averages is detrimental. Intrabreed craniofacial variation is detectable and, in some instances, notably diverse such as in the Labrador Retriever and Yorkshire Terrier (Chapter 3). This intrabreed variation is well documented and has even been the fundamental principle upon which the size-related *IGF1* locus was identified within the morphologically variable Portuguese Waterdog (Chase et al. 2002). Averaging the diverse craniofacial variation within breeds at a time in which one is attempting to determine the causes of this variation would undermine such the purpose of the study. Finally, the use of uncorrected morphological measurements can be confounded by the influence of allometry. Previous studies have begun to address this and methods upon which best to remove the influence of allometry were discussed in Chapter 3.

Throughout the previous studies, individual-level genotypes have been separated from phenotypes. Even in the most recent and most successful analysis of canine craniofacial variation by Schoenebeck et al., used disconnected phenotypes and genotypes (Schoenebeck *et al.* 2012). The morphometrics of archived skeletal remains were analysed and used as surrogate phenotypes with DNA collected from modern-day extant show dogs. Evidently, this method was successful in identifying a variant in *BMP3* to be associated with small brachycephalic breeds however in order to exploit the individual-level geometric morphometric approach to phenotyping, a corresponding source of DNA and subsequent microarray data must be acquired. Furthermore, acquisition of individual-level data permits the utilization of mixed

breed dogs – something that has never been adopted before. This inclusion of mixed breed dogs was hoped to break up the extensive linkage disequilibrium that encompasses selective sweeps across breeds to assist in the fine mapping for QTL and to help predict the proportion of variance explained by QTL.

Finally, previous studies have assumed breeds sharing a physical characteristic do so because the underlying genetic drivers are shared by virtue of sharing alleles identical-by-descent (IBD). On the whole, this is likely to be true however it is not absolute. Exceptions include the presence of the *IGF1* small haplotype in giant breed such as the Mastiff and Bernese Mountain Dog (Sutter *et al.* 2007), the *BMP3* brachycephalic missense mutation in the non-brachycephalic Scottish terrier (Schoenebeck *et al.* 2012) and the lack of the *FGF5* long coat variant in the Afghan Hound (Housley and Venta 2006; Dierks *et al.* 2013). Such exceptions should be identified and tolerated during fine mapping, something that was not previously done when mapping the CFA1 brachycephaly locus (Bannasch *et al.* 2010).

To this end, this chapter aimed to resolve the CFA1 QTL for canine brachycephaly through the adoption of modified techniques from previous studies. Precisely, this included the use of a previously acquired individual-level high-definition non-allometric phenotype for craniofacial shape variation as an input for a genome-wide association analysis. Individual-level haplotypes were phased and used to define the CFA1 critical interval within which variants for brachycephalic breeds were called and filtered against non-brachycephalic dog's that did not carry the associated haplotype. These same methods were then adopted to further contribute to and define the genetic loci associated canine body size, or in this instance, neurocranium centroid size.

4.2 Materials and Methods

4.2.1 Participants

The same participants and study cohorts that were used to address morphological variation in Chapter 3 were used here. Briefly, 374 dogs were included

in the first study cohort which was subsequently expanded to 560 dogs. Every participant described here has a matching genotype and phenotype which is a distinction from previous studies that use surrogate phenotypes. Throughout this chapter, the use of study cohort one or two will be explicitly listed in text.

4.2.2 Microarray Genotype Preparation

gDNA for all participants in this study was extracted following the protocols described previously (see Chapter 2). gDNA quality and quantity was determined using the NanoDrop device and PicoGreen™ assay, respectively (see Sections 2.4.1 and 2.5.3). All samples contained high quality gDNA with 260/280 ratios greater than 1.6 and 260/230 ratios greater than 2.0. Samples were arrayed onto batches of 96-well plate provided by Edinburgh Genomics, UK to give a final concentration of 50 ng/µL once rehydrated in 50 µL deionised water following the use of a speed vac (see Section 2.5.6.1). Prepared plates were submitted to Edinburgh Genomics to be genotyped on the Illumina 170,000 marker CanineHD Whole-Genome Genotyping BeadChip. Genotype calls were mapped to CanFam3.1 coordinates (Broad, September 2011). PLINK (v1.07) was used to remove SNPs with minor allele frequencies (MAF) < 0.05 and individuals with > 0.1 missing markers (Purcell *et al.* 2007). Genotypes were phased with SHAPEIT (v2.r837) using default parameters that include 500 states (Delaneau *et al.* 2012). Imputations were done with Minimac2 (2014.9.15) using 40 rounds and 1,000 states (Fuchsberger *et al.* 2014). Post-processing by fcGene (v1.0.7) removed genotypes with R2 < 0.3 and MAF < 0.05 (Roshyara and Scholz 2014). In total, 139,260 SNPs remained for analysis.

4.2.3 Genomic Structure Analysis

The population structure of the first study cohort was assessed using STRUCTURE (v2.3.4) (Pritchard *et al.* 2000). SNP microarray data was first pruned to only keep markers with predicted linkage disequilibrium (R^2 <0.5) using PLINK (v1.9) (Chang *et al.* 2015). Subsequently the pruned dataset was reformatted using the "--recode structure" argument in PLINK (v1.9) to generate an appropriate input file. Default parameters were used to run STRUCTURE and the number of populations (K)

were varied to reflect the expected number of breeds. fastSTRUCTURE (v1.0) was used to assess the genomic structure of the second cohort (Raj *et al.* 2014). No data modification was required prior to the use of fastSTRUCTURE. Default parameters were used and the number of populations were varied to include 2, 3, 4 and 92 populations. A maximum K value of 92 was used to reflect the 92 breeds in the study – each breed would be expected to have its own population membership. Graphics were generated using Microsoft Excel and modified in Adobe Photoshop CC 2015.

4.2.4 Genome-wide Association Analysis

The Genome-wide Efficient Mixed Model Association (GEMMA) (v0.94.1) was used to perform univariate linear mixed models for genome-wide association analyses (Zhou and Stephens 2012). During the analysis, a kinship matrix was generated using default parameters and subsequently incorporated to account for population structure. Covariates were also used in the analysis and varied between studies to further remove effects of population structure. The first study cohort used sex and up to ten genotype principal components (PC) to account for population structure. The genotypic PCs were generated using the "--pca" argument in PLINK (v1.9) and increasing numbers of PCs were applied as covariates to the association studies until genomic inflation was managed. To determine the level of genomic inflation, QQ-plots were plotted to visualise the relationship between expected and observed distributions of p-values which were quantified using the genomic inflation factor, lambda (λ) (Yang *et al.* 2011). The p-values of each association analysis (with different numbers of PC covariates) were sequentially input to Equation 4.1 until genomic inflation was reduced ($\lambda \approx 1$). A total of ten PCs were used for the neurocranium centroid size association analysis whilst five PCs were used for the viscerocranium and mandible analyses. To confirm that any remaining SNPs that deviate from the abline in the QQ-plot are the result of significant associations, index SNPs and markers in high linkage disequilibrium ($r^2 > 0.2$) were pruned from the dataset, association analyses re-run and QQ-plots re-plotted.

In the second study cohort, the covariates were changed to include sex, age, CT image slice thickness (1 or 2 mm slices) and Hounsfield units used in the reconstruction of skull isosurfaces. Lambda scores and QQ-plots indicated minimal genomic inflation which negated the use of PCs as covariates in this analysis. A highly stringent Bonferroni correction was used to set the threshold for genome-wide significance (Equation 4.2). The number of markers used in the first and second study cohorts were 139,260 and 152,111 giving -log*P* value thresholds of 6.44 and 6.48, respectively.

$$\lambda = median \frac{qchisq(1-p,1)}{qchisq(0.5,1)}$$

Equation 4.1 – Genomic Inflation Equation. The genomic inflation factor (λ) is calculated using the above equation where *p* are the observed p-values for an association test and *qchisq* is the non-central Chi-squared distribution function available in RStudio (v1.0.153).

$$P = -log_{10}(\frac{0.05}{Total\ Markers})$$

Equation 4.2 – Bonferroni Correction Equation. The stringent Bonferroni correction threshold (-log*P*) for genome-wide significance is calculated by the above equation where the *Total Markers* are the number of markers used in the association analysis.

4.2.5 Fine Mapping

Manhattan plots focussing on each individual QTL and spanned the region of interest –1 Mb up- and downstream of the genome-wide significant markers. Linkage disequilibrium between each index marker and their surrounding 500 flanking markers were calculated using the PLINK (v1.9) LD statistic argument, "--r2". Linkage disequilibrium values (r^2) were used with the raw output from the GEMMA association analysis to generate zoomed Manhattan plots using a custom R script.

In order to define the critical interval on CFA1, data was used from the first study cohort. Haplotype association analyses were used to refine the region of

interest surrounding significant markers. For a given phenotype, the PLINK (v1.07) "--hap-logistic" argument was used to with 3, 5 and 10 marker sliding windows. The analysis was carried out for ~1 Mb in both directions from the index marker. The raw text output file which includes p-values and the positions of the first and last SNPs in a haplotype were used to generate plots in RStudio. The haplotype with the greatest association and ~500 Kb flanking regions were used to define the region of interest.

Individual-level haplotypes were assessed to define the critical interval. Genotypes were phased with SHAPEIT (v2.r837) (Delaneau *et al.* 2013). For the CFA1 locus, CFA1 haplotypes for the region of interest spanning 55,628,389 – 56,503,947 were ordered by viscerocranium PC1 with the most brachycephalic dogs at the top and lesser so at the bottom. Alleles were contrasted to the most frequent haplotype in the brachycephalic cohort (viscerocranium PC1 < -0.2). Within the viscerocranium PC1 < -0.2 subgroup, the critical interval boundaries were defined by a minimum of 3 or more meiotic recombination events on either side.

Critical interval boundaries established for all other QTL (including CFA5 for viscerocranium PC1 and all neurocranium associated QTL) were defined using data derived from the second study cohort. Haplotype association analyses were not used. For each of these QTL, methods were adapted from those described for CFA1. Firstly, haplotypes at QTL associated with the neurocranium centroid size were ranked by neurocranium centroid size (rather than viscerocranium PC1) from smallest to largest. Secondly, the origin of the most frequent haplotype to be used as a consensus sequence was modified (Table 4.1). For high frequency positions which were shared across many breeds as indicated by the distribution of the derived index markers, the smallest quartile of dogs (neurocranium centroid size < 126) were pooled together. Within this subgroup, the most common haplotype was used as a consensus sequence. For less frequent positions which were only present in a subset of breeds, the most common haplotype was found in a subset of breeds enriched with the derived marker allele. For example, the derived marker allele at the CFA5 locus (chr5:33,750,479) was observed to be enriched in the American Bulldog, Boxer,

Bulldog, French Bulldog and French Mastiff. These breeds were grouped and the most frequent haplotype used as a consensus. This strategy was adopted to prevent alternative haplotypes at high frequencies which are not driving the genomic associations from introducing erroneous consensus haplotypes. In this example, the highly populous Pug, which was not observed to carry the derived marker allele would have introduced a high frequency haplotype to use as a consensus. This would have made it impossible to define a critical interval once all other alleles were contrasted to this false consensus. A full list of methods to defining population subgroups for consensus haplotype assignment is given in Table 4.1. Once the

Phenotype	Index Marker	Consensus Source
Viscerocranium	chr1:55,983,871	Viscerocranium PC1 < -0.2
	chr5:33,750,479	American Bulldog, Boxer, Bulldog, French Bulldog, French Mastiff
Neurocranium	chr3:91,103,945	Smallest quartile Neurocranium centroid size (<126)
	chr4:67,109,434	Pug, Border Terrier, Bichon Frise, Border Collie, Chihuahua, Cairn Terrier, Dachshund, Maltese Terrier, Miniature Schnauzer, West Highland White Terrier
	chr7:30,243,851	Chihuahua, Border Terrier, Dachshund, Lhasa Apso, Maltese Terrier, Manchester Terrier, Papillion, Pekingese, Shih Tzu
	chr7:43,719,549	Smallest quartile Neurocranium centroid size (<126)
	chr10:8,183,593	Smallest quartile Neurocranium centroid size (<126)
	chr15:41,221,438	Smallest quartile Neurocranium centroid size (<126)
	chr34:18,540,901	Border Terrier, Boston Terrier, Cairn Terrier, French Bulldog, Jack Russel Terrier, Maltese Terrier, Miniature Schnauzer, West Highland White Terrier

Table 4.1 – Determining the Consensus Haplotype. For each phenotype QTL, the haplotypes are pooled together into an appropriate subgroup to identify the most common derived haplotype to contrast all other alleles to. For common index alleles, the extremes of the phenotypic continuum can be used. For rarer alleles, breeds that are enriched for the index allele are isolated to increase the likelihood of accurately capturing the derived haplotype.

consensus haplotype was identified it was used to contrast all other alleles. Critical interval boundaries were defined by 3 or more meiotic recombinations as done previously.

4.2.6 Library Preparation and Sequencing

All gDNA destined for whole-genome sequencing was quality checked using the Agilent TapeStation and surpassed DNA integrity numbers (DIN) of 8.0 (see Section 2.4.4). Sequencing services were provided by Edinburgh Genomics, UK which in brief, included the following. DNA libraries were prepared using either SeqLab TruSeq Nano DNA library prep HT or Illumina TruSeq DNA nano DNA library kits. Eight brachycephalic dogs were resequenced on an Illumina HiSeq 2000 to approximately 14-33X depth. Paired-end libraries had an average insert size of 550 bp and read length of 125 bp. A further thirty dogs were resequenced using the Illumina HiSeq X platform to > 40X depth with the paired-end library having an average insert size of 450 bp and a 150 bp read length. DNA short-reads were aligned to the CanFam3.1 reference sequence (Broad, September 2011) using the Burrows-Wheeler Aligner (BWA)-MEM software (v0.7.16a-r1181) (Li 2013).

4.2.7 Variant Analysis (CFA1 Only)

SNPs and small insertion/deletion (INDEL) variants were called within the CFA1 critical interval (55,850,299-56,037,676) for the sequenced dogs using the Genomic Analysis Toolkit (GATK) (v3.7) (McKenna *et al*. 2010; DePristo *et al.* 2011; Van der Auwera *et al.* 2014). Variants were compared against those from four ancestral wolves in the Dog Genome SNP Database (DoGSD) and dogs in the Dog Biomedical Variant Calling Database (DBVCD). The DBVCD is a collection of canine whole genome sequences which allows contributors to share variant calls across the collection. At the time of use, the consortium was composed of 304 individuals representing modern dog breeds and a Basenji which have diverse range craniofacial morphologies. SNP and small INDEL variants across all individuals in this study were hard filtered using SnpSift (v4.0) using the criteria is Table 4.2 (Cingolani *et al.* 2012). For a variant to be retained for further consideration, it must satisfy all of these

criteria. The deep coverage and large insert sizes of the thirty-eight resequenced dogs permitted for the detection and filtering of large structural variants using Pindel (v0.2.3) (Ye *et al.* 2009). Structural variants were hard filtered using the criteria in Table 4.2. The filtering criteria for both SNP/INDEL and large structural variants were established by the presence or absence of the brachycephaly-associated haplotype across brachycephaly and dolichocephalic dogs with confirmed skull shapes. During variant filtering several logical assumptions were made. Firstly, brachycephalic dogs carrying the associated haplotype would be homozygous (or heterozygous) for the causal mutation(s) within the critical interval. Second, the sharing of a common haplotype would indicate an identity-by-descent inheritance of the same causal

Variant	Breed	Filtering Criteria
SNP/INDEL	Basenji	Homozygous Ancestral
	Wolf	
	Wolf	
	Wolf	
	Wolf	
	Pug	Homozygous Derived
	American Bulldog	
	French Bulldog	
	French Bulldog	
	French Bulldog	
	Saluki	Homozygous Ancestral
	Whippet	
	Italian Greyhound	
	Airedale Terrier	
Structural	Basenji	Homozygous Ancestral
	Wolf	
	Wolf	
	Wolf	
	Wolf	
	Pug	Homozygous Derived
	American Bulldog	
	French Bulldog	

Table 4.2 – Filtering criteria for CFA1. For each variant type to remain after filtering, they must satisfy all of the above criteria for each of the breeds (or wolf).

mutation(s) from a common ancestor. Third, the CanFam3.1 dog genome assembly is derived from a brachycephalic Boxer (which is homozygous for the associated haplotype) and so it is likely a carrier of, or fixed for the causal mutation(s) and so the variant would appear as a reference allele. Fourth, the causal mutation(s) are derived and therefore absent from the wild candid population, including the dog's ancestor, the gray wolf. Lastly, dolichocephalic dogs without the associated haplotype cannot carry the causal mutation(s).

4.3 Results

4.3.1 Genomic Structure

In addition to assisting QTL discovering and fine mapping of these loci, SNP microarray data can also be used to infer global trends of individuals at a genomic level. This population structure can prove to be detrimental during certain analyses such as a genome-wide association analyses but can be accounted for through the inclusion of Kinship matrices during testing (see Section 4.2.4). Here, fastSTRUCTURE has been used to visualise the population structure of all 560 dogs arrayed across both cohort one and two (Figure 4.1). During testing, the percentage membership of a dog's genomic structure is assigned to a pre-set number of populations (K). When K = 2, the Pugs are identified as a distinct breed from others – this is likely due to the high representation of the breed in the test and the distinct homogeneity of the breed compared to more populous breeds in the study such as the Labrador retriever (Figure 4.1D). When K = 3, a broad collection of dog breeds begin to show membership to this population (Figure 4.1E). This includes breeds such as the Boxer, American Bulldog, Bulldog, Bull Mastiff, French Bulldog, Dogue de Bordeaux, Pitbull Terrier and Staffordshire Bull Terrier. Broadly, these breeds are representative of the Molosser category of dogs which are larger, well-muscled breeds (Parker *et al.* 2004). Sequential increases in population number continue to identify broad population structures, for example when K = 4, the retriever class begins to emerge which includes the Flat Coat Retriever, Golden Retriever and Labrador Retriever (data not shown).

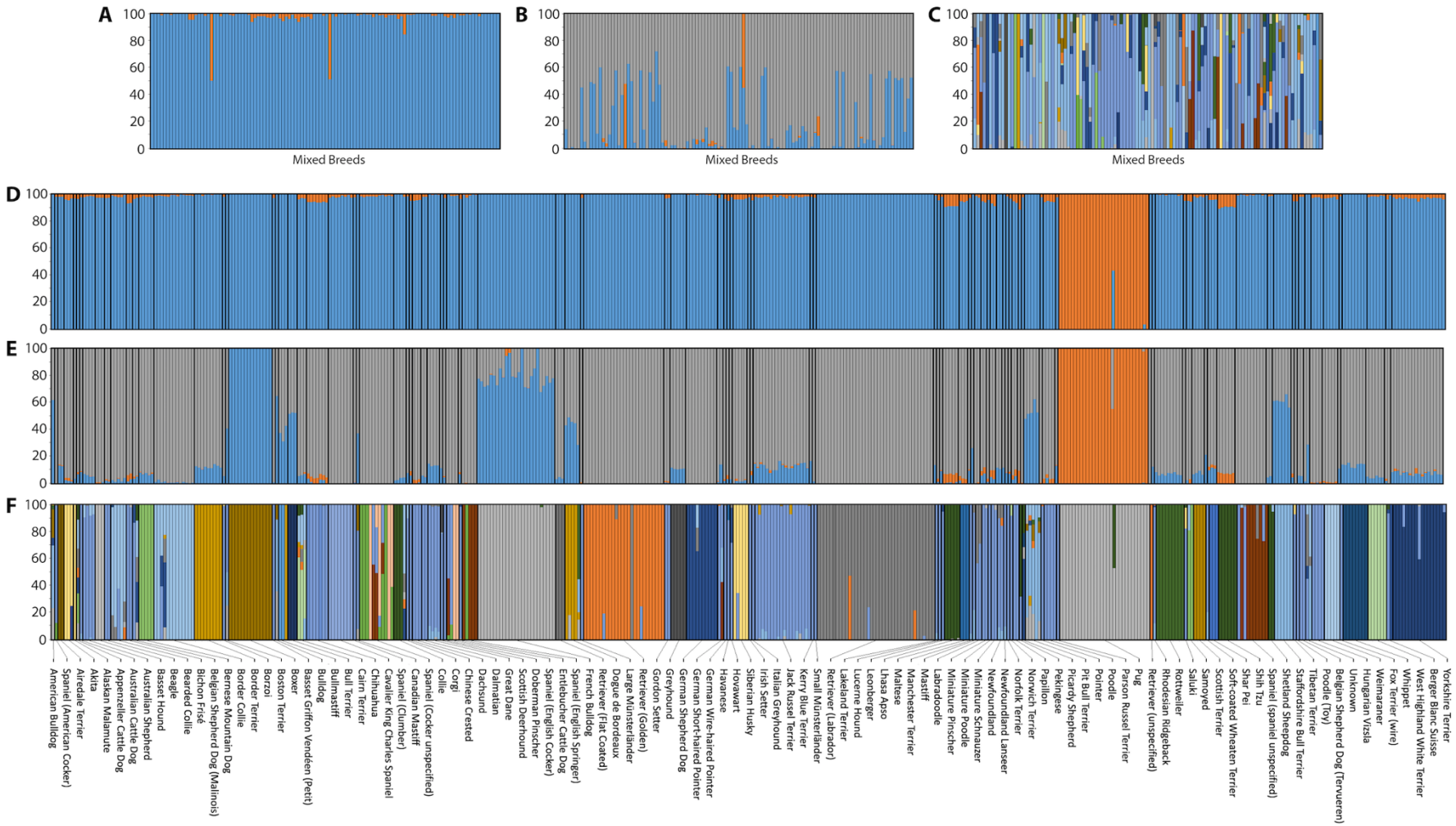


Figure 4.1 – Population Structure of all Study Dogs. Each individual dog is represented by a single vertical bar separated by black lines with (A-C) mixed-breed and (D-F) owner-defined breeds being grouped together by broad black lines. The percentage assignment (y-axis) of each individual to a population cluster (K) is indicated by the coloured bars. The number of population clusters are (A,D) 2, (B,E) 3 and (C,F) 92.

Increasing the population number to the predicted number of breeds (K = 92) can indicate breed-level structures (Figure 4.1F). Numerous breeds such as the Bearded Collie, Chihuahua, Greyhound and Shih Tzu are assigned entirely to a single population indicating that individuals within these breeds are very alike at a genomic-level and so likely represent dogs which closely adhere to the breed standard. Within the Pug population, one individual dog is distinct from the others as it has equal membership to the Pug population and another – the Shih Tzu (Figure 4.1F). This likely represents a first-generation offspring between a Pug and a Shih Tzu – a breeding combination increasing in popularity to generate a "Pug Zu". Furthermore, a dog labelled as a Labrador Retriever in fact has ~100% membership to the Flat Coat Retriever population suggesting that the clinical records for this dog are incorrect. Similar approaches to investigating potential breed composition can be applied to the mixed-breed dogs (Figures 4.1A, B and C). A notable proportion of mixed-breeds dogs have shared genomic structure with the purebred Staffordshire Bull Terrier. This observation is likely caused by the high number of Staffordshire bull terrier-like stray or abandoned dogs taken in by the Edinburgh Dog and Cat Home which visit the Hospital for Small Animals and formed part of this study cohort.

Breeds with more diverse genomic structures include the Cairn terrier and Pitbull Terrier. It is unclear whether this observation is reflected across these entire breeds or if it's restricted to a small population of individuals under less stringent breeding regulations that have been included in this study.

4.3.2 Study Cohort 1

In Chapter 3, geometric morphometric analyses of the canine skull were performed on 374 dogs which constituted study cohort 1. Here, the principal components (PC) of the whole skull, viscerocranium and mandible were used as input phenotypes for genome-wide association studies (GWAS). Importantly, each dog in the study that received CT imaging and subsequent geometric morphometric analysis also supplied a source of DNA for SNP microarray genotyping. This allowed for individually matched phenotypes and genotypes which negates the use of surrogate

or breed-average phenotype-genotype matches which have been adopted in previous studies (Jones *et al.* 2008; Bannasch *et al.* 2010; Boyko *et al.* 2010; Schoenebeck *et al.* 2012).

4.3.2.1 Genome-wide Association Analysis

An integral component of the current study design is the use of the neurocranium centroid size as a proxy for skeletal size and its subsequent use in removing the influence of allometry in this data. This body size proxy was necessary as patient weights (or other body size measurements) were not available for every study dog. To demonstrate that the neurocranium centroid size accurately reflects the skeletal size in the dog, a GWAS was performed using the centroid size as a phenotype (Figure 4.2A). In total, 32 markers were associated with the neurocranium centroid size which surpassed a stringent threshold for genome-wide significance set by a Bonferroni correction. The markers were distributed across five genomic loci on five different chromosomes (Table 4.3). Index markers for the QTL on canine chromosome (CFA) 7, 10 and 15 resided within introns of the SMAD family member 2 (*SMAD2*), high-mobility group AT-hook 2 (*HMGA2*) and insulin-like growth factor 1 (*IGF1*) genes respectively (Table 4.3). The index marker on CFA3 is intergenic and situated ~72 kb and ~176 kb upstream from the ligand-dependent nuclear receptor corepressor-like (*LCORL*) and condensin complex subunit 3 (*NCAPG*) genes respectively. The anaphase promoting complex subunit 13 (*ANAPC13*) is also downstream of the index marker. Each of the *SMAD2*, *HMGA2*, *IGF1* and *NCAPG*/*LCORL* loci have been identified previously to be associated with body size variation in the dog and across species which confirms that the neurocranium centroid size is an appropriate proxy (Chase *et al.* 2002; Sutter *et al.* 2007; Boyko *et al.* 2010; Vaysse *et al.* 2011; Rimbault *et al.* 2013; Hayward *et al.* 2016).

In a distinction from previous body size studies in the dog, here a signal for the fibroblast growth factor 4 (*FGF4*) retrogene on CFA18 has been detected with neurocranium centroid size. This is a surprising discovery as the *FGF4* retrogene was first identified to correlate with limb shortening in breeds such as the Dachshund and

Basset hound. This current data suggests it may also have a role in regulating skull size too.

The Manhattan plot for non-allometric viscerocranium PC1 revealed a single locus on CFA1 to be comprised of 3 markers surpassing genome-wide significance (Figure 4.2B). The index marker for this locus (chr1:55,983,871) is situated within an intron of the SPARC-related modular calcium-binding protein 2 (*SMOC2*) gene (Table

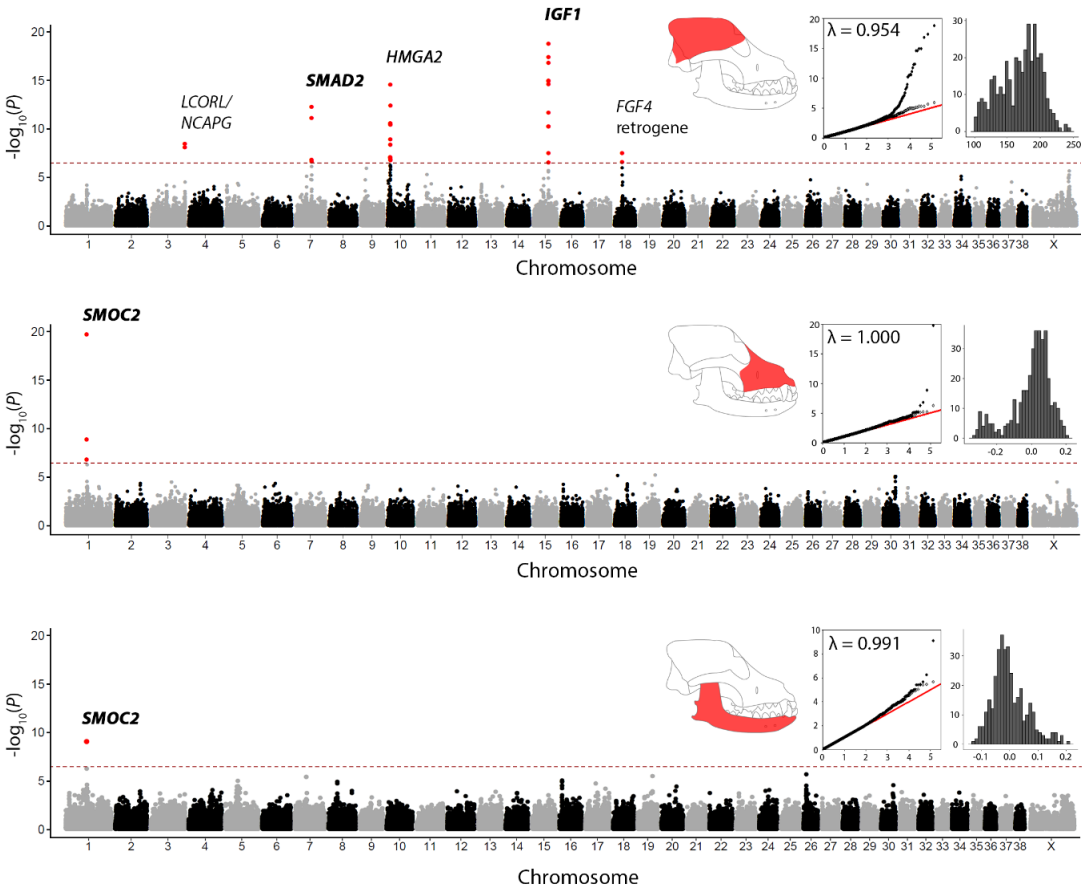


Figure 4.2 – Study Cohort 1 GWAS. Manhattan plots for the (A) neurocranium centroid size, (B) viscerocranium PC1 and (C) mandible PC1. Markers surpassing the threshold for genome-wide significance as calculated by a Bonferroni correction are coloured red. Positional gene candidates are given above each QTL with intragenic markers in bold text. Insets: anatomical substructures analysed are highlighted in red. QQ-plots for the predicted (x-axis) and observed (y-axis) -log$_{10}$[p-values] for association tests. P-values plotted using the entire marker dataset (black circles) and pruned dataset following the removal of significant markers and markers in high linkage disequilibrium (r^2>0.2). Abline (y=x) plotted in red. Genomic inflation value (λ) shows minimal inflation across analyses. Distribution plots displaying the frequency (y-axis) of the phenotype scores (x-axis).

4.3). The same index marker was also significantly associated with non-allometric mandible PC1. This was the only marker to reach genome-wide significance for the mandible dataset.

Across all GWAS, the genomic inflation factor, lambda (λ) is approximately 1 (Figure 4.2). Lambda values of 1 indicate no genomic inflation. QQ-plots for each association test confirm that observed p-values adhere to the abline and only deviate

Dataset	Position	Candidate Gene	Allele	P-value	beta
V1	1:55,983,871	*SMOC2*	G > A	1.91×10^{-20}	-0.07
M1	1:55,983,871	*SMOC2*	G > A	8.43×10^{-10}	0.02
N1	3:91,103,945	*LCORL/NCAPG*	G > T	3.64×10^{-9}	10.22
	7:43,719,549	*SMAD2*	A > G	5.71×10^{-13}	-11.15
	10:8,183,593	*HMGA2*	C > T	3.06×10^{-15}	-14.06
	15:41,257,020	*IGF1*	C > T	1.73×10^{-19}	-16.34
	18:20,272,961	*FGF4* retrogene	A > C	3.31×10^{-8}	-11.41
V2	1:55,983,871	*SMOC2*	G > A	6.67×10^{-39}	-0.07
	5:33,750,479	*NTRN1*	G > A	2.26×10^{-9}	-0.04
N2	3:91,103,945	*LCORL/NCAPG*	G > T	1.12×10^{-9}	8.99
	4:67,109,434	*GHR*	T > C	6.08×10^{-8}	-5.07
	7:30,243,851	*TBX19*	A > G	3.19×10^{-7}	-5.32
	7:43,719,549	*SMAD2*	A > G	3.95×10^{-14}	-9.44
	10:8,183,593	*HMGA2*	C > T	3.01×10^{-22}	-14.37
	15:41,221,438	*IGF1*	A > G	4.39×10^{-28}	11.36
	18:20,310,833	*FGF4* retrogene	T > G	2.26×10^{-12}	-9.75
	34:18,540,901	*IGF2BP2*	A > C	1.15×10^{-7}	-6.79
	X:103,307,510	-	C > T	9.12×10^{-9}	-6.81

Table 4.3 – Index Markers for Study Cohorts 1 and 2. Index markers discovered in the viscerocranium PC1 (V), mandible PC1 (M) and neurocranium centroid size (N) association tests for study cohorts 1 and 2 (cohort given in brackets). Positional gene candidates are highlighted in bold when the index marker is intragenic. Alleles are given as reference > alternate. P-value of the index marker generated in the linear mixed model analysis.

to the right of the plot (black circles). This deviation largely disappears once the test is re-run having removed the markers surpassing genome-wide significance and any markers in linkage disequilibrium ($r^2 > 0.2$) with them (grey circles). These tests conclude that there is minimal genomic inflation in this data, and thus no evidence of population substructure.

4.3.2.2 Defining the CFA1 Critical Interval

The objective of the first study cohort was to identify and resolve QTL associated with canine brachycephaly. The QTL on CFA1 represents a broad selective sweep that has been identified in several previous studies (Bannasch *et al.* 2010; Boyko *et al.* 2010; Schoenebeck *et al.* 2012). In the current analysis, the selective sweep was tagged by three markers which surpassed genome-wide significance. A further fourteen non-significant markers were in high linkage disequilibrium ($r^2 > 0.2$) with the index marker (Figure 4.3A). To begin to refine this selective sweep, a haplotype association analysis using a 10-marker sliding window was performed for markers within ~1 Mb of the significant markers (Figure 4.3B). This revealed a collection of highly associated haplotypes within the window chr1:55,881,672-56,020,217. Genotypes within this interval including ~500 kb flanking regions were phased. Since the viscerocranium PC1 distribution is bimodal (see insert of Figure 4.2B) and that the majority of breeds within the smaller peak (PC1 < -0.20) are brachycephalic, an assumption was made that the causal variant and therefore associated haplotype would be enriched within subgroup. To this end, the most frequent haplotype across individuals with a viscerocranium PC1 < -0.20 was used as the consensus haplotype that all others would be contrasted to. The resulting haplotypes were plotted with alleles matching the consensus in light grey and alternate alleles in dark grey (Figure 4.3C). Haplotypes were ordered by viscerocranium PC1 score to prioritise the most brachycephalic dogs at the top and the lesser-so at the bottom. Boundaries of the haplotypes were defined by a minimum of three meiotic recombinations on either side.

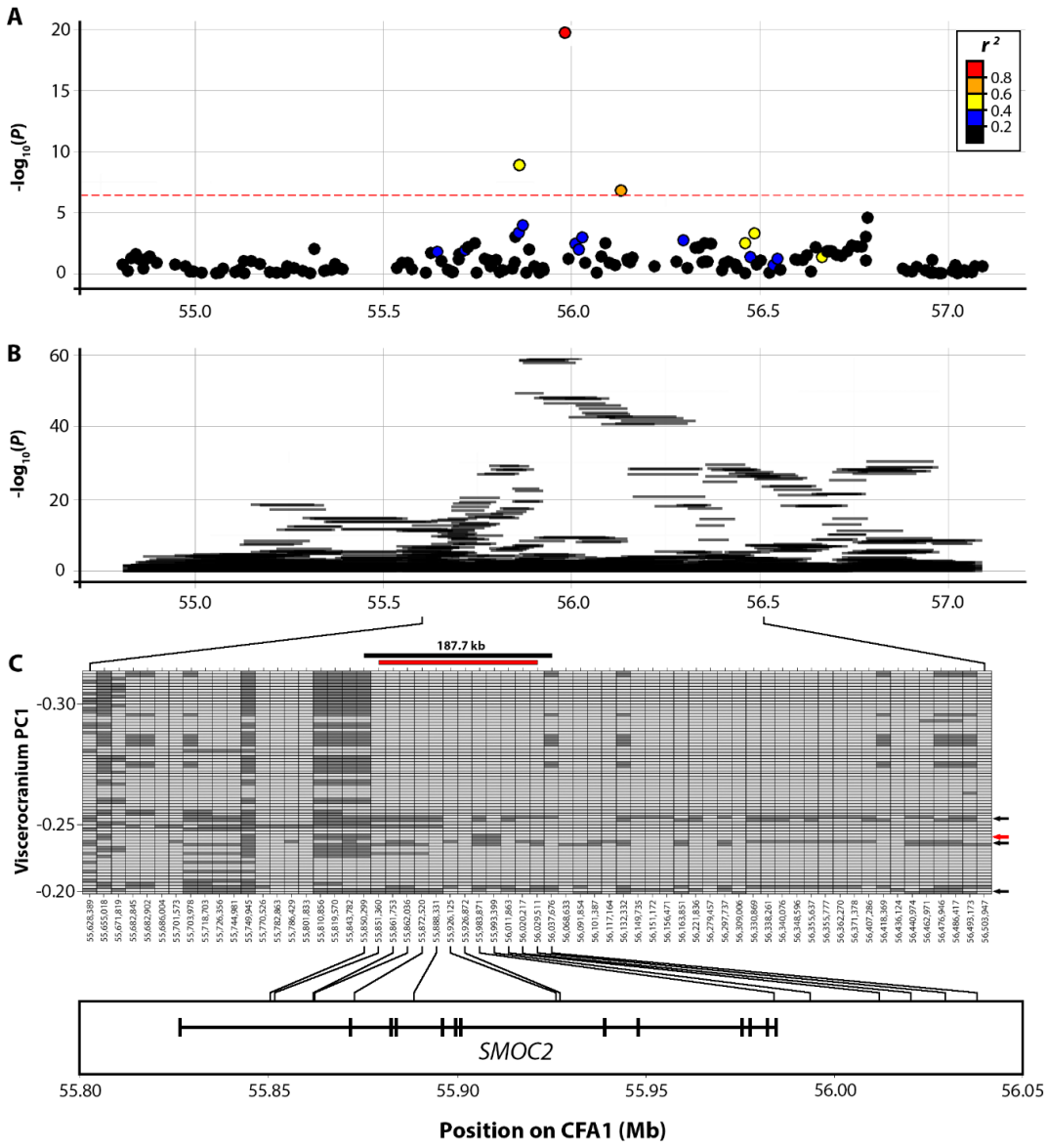


Figure 4.3 – Defining the CFA1 Critical Interval. (A) Manhattan plot of the region surrounding the three markers surpassing genome-wide significance for viscerocranium PC1. Markers in linkage disequilibrium (r>0.2) with the index marker (chr1:55,983,871) are coloured. (B) Ten-marker sliding window haplotype association test for the same interval. A cluster of highly significant haplotypes ($p = 9.95 \times 10^{-60}$) spanning CFA1 from 55,881,672 to 56,020,217 are associated with viscerocranium PC1. (C) Genotypes within this region and ~500 kb of flanking region were phased. The most frequent haplotype within this interval for the brachycephalic cohort (viscerocranium PC1 < -0.20) was used as a consensus haplotype that all others were contrasted to. Matching alleles are coloured light grey whilst alternate alleles coloured dark grey. Haplotype boundaries are defined by 3 meiotic recombinations. A 12 marker haplotype (red bar) flanked by a 187.7 kb critical interval (black bar) spanning chr1:55,850,299-56,037,676 was defined. Black arrows denote dogs that do not carry the associated haplotype. Red arrow denotes a chromosome with a possible novel recombination event. The critical interval spans the SMOC2 gene.

A twelve-marker haplotype (red bar) was defined by the 187.4 kb critical interval (black bar) which extended between chr1:55,850,299-56,037,676 (Figure 4.3C). This haplotype was overrepresented amongst the brachycephalic cohort (viscerocranium PC1 < -0.20) and was present in 63 of 74 (85.1%) chromosomes. Conversely, the haplotype was observed amongst just 28 of 674 (4.2%) chromosomes of mesocephalic to dolichocephalic skull shapes (viscerocranium PC1 > -0.20). Suggestive of an effect, a student's t-test revealed a highly significant difference between the viscerocranium PC1 scores between carriers and non-carriers ($p = 4.86 \times 10^{-49}$). Interestingly, the associated haplotype was completely absent (homozygous ancestral) from three individuals that were classified as brachycephalic, albeit less-extreme brachycephalic dogs (Figure 4.3C). This included two Dogue de Bordeaux and a Chihuahua (black arrows). The fastSTRUCTURE analysis of these dogs (Figure 4.1F) revealed a higher degree of admixture between these individuals than others in their designated breeds indicating that they may in fact be partly outbred. A Japanese Chin did not have the full-length 12 marker haplotype (red arrow) however it is unclear whether this may be due to a recent meiotic recombination event that hasn't accumulated sufficiently across dogs or if in fact, this chromosome of the Japanese Chin does not carry the causal mutation.

Eight of the twelve marker haplotype are situated within introns of the *SMOC2* gene whilst the remaining four markers span up to ~43 kb downstream of the gene (Figure 4.3C). No other genes are annotated within the critical interval.

4.3.2.3 Variant Calling and Filtering

To identify causal variants within the CFA1 critical interval, eight brachycephalic dogs including an American Bulldog, Boxer, Pug, Cavalier King Charles Spaniel, Affenpinscher, Bull Mastiff, French Bulldog and French Mastiff were whole genome sequenced. A further thirty non-brachycephalic breeds were also sequenced. SNPs and small INDELS were called across these dogs and compared against those in the

ancestral gray wolf and wider DBVCD consortium and filtered on the presence of the associated haplotype (see Section 4.2.7).

Within the 187,377 bp critical interval on chromosome one, 3,674 SNPs and small indels were called (Table 4.4). After hard filtering, using the criteria in Section 4.2.7, four SNPs situated within introns 1, 7 (x2) and 12 of *SMOC2* remained. Structural variants were also called within the same 187,377 bp critical interval. Before filtering, a total of 162 structural elements were called which were filtered to a single ~1,500 bp insertion within intron 8 of *SMOC2*. This insertion is a fragment of a long interspersed nuclear element type 1 (LINE-1).

The biological significance of the five remaining putatively causal variants within *SMOC2* are explored in Chapter 5. This includes the embryological localisation of mouse and chicken *Smoc2*, the effects of the variants on transcriptomic profiles and modelling of the phenotypic effects.

Software	GATK/SnpSift	Pindel
Variant type	SNPs/INDELS	Structural variants
Bases pairs analysed	187,377	187,377
Pre-filtering	3,674	162
Post-filtering	4	1

Table 4.4 – Variant Filtering Summary. SNPs/INDELS and structural variants were called within the same CFA1 critical interval (chr1:55,850,299-56,037,676) for canine brachycephaly. The total variants called were filtered using criteria in Table 4.2. Post-filtering variants have passed the filtering criteria and are putatively causal.

4.3.3 Study Cohort 2

The continued recruitment of participants has permitted for the expansion of the study from 374 dogs in cohort 1 to 560 in cohort 2. Coupled with modifications to the effectively removing the allometric component to the geometric morphometric data (see Section 3.3.3.1), the increase in participants and diversification of breeds available have increased the power of this study.

4.3.3.1 Genome-wide Association Analysis

Previously, 32 markers distributed across 5 QTL surpassed genome-wide significance for the neurocranium centroid size (Figure 4.2). This included the positional candidate genes *SMAD2*, *HMGA2*, *IGF1, LCORL*/*NCAPG* locus and the *FGF4* retrogene. These same QTL were identified in the current analysis in addition to a further 4 novel QTL for this study positioned on CFA4, 7, 34 and X (Figure 4.4A). The index markers on CFA4 and CFAX are positioned within introns of the growth hormone receptor (*GHR*) and muscleblind-like splicing regulator 3 (*MBNL3*) genes respectively (Table 4.3). Meanwhile the index marker on CFA7 (chr7:30,243,851) is positioned ~10 kb downstream of the T-box transcription factor 19 (*TBX19*) gene whilst the closest gene to the CFA34 marker (chr34:18,540,901) is insulin-like growth factor 2 mRNA-binding protein 2 (*IGF2BP2*).

These data continue to demonstrate that the neurocranium centroid size effectively acts as a proxy to body size by identifying previous body size-associated loci. Each of the QTL on CFA4, 7 and 34 have been identified previously (Boyko *et al.* 2010; Rimbault *et al.* 2013; Hayward *et al.* 2016). Several positional candidate genes have been alluded to on the X chromosome however none have identified the *MBNL3* gene.

With regards to the association analysis of viscerocranium PC1, the CFA1 *SMOC2* locus was identified again (Figure 4.4B). Here, the same index marker at position chr1:55,983,871 was identified as previously described however the total of markers surpassing genome-wide significance increased from 3 to 5 (Table 4.3). In addition to this locus, a further 5 markers surpassed genome-wide significance at a QTL on CFA5. The index marker for this locus is positioned within an intron of the netrin 1 (*NTRN1*) gene. Previous studies investigating the brachycephalic phenotype also discovered a QTL on CFA5 however those loci were positioned ~4 Mb upstream from the locus described here (Boyko *et al.* 2010; Schoenebeck *et al.* 2012).

Suggestive peaks begin to emerge for the viscerocranial dataset and includes CFA9, 18, 24 and 30 (Figure 4.4B). Here, suggestive peaks are QTL that have multiple

markers in close proximity that are approaching, but don't reach the Bonferroni threshold. These peaks tend to be distinctive from the surrounding markers in the Manhattan plots and are identified purely by visual inspection of the plots. Index markers for each of these QTL reside in introns of interesting positional candidate genes. This includes histone deacetylase 5 (*HDAC5*, chr9:19,267,017), zinc finger protein GLI3 (*GLI3*, chr18:7,835,560), agouti-signalling protein (*ASIP*, chr24: 23,386,821) and coronin 2B (*CORO2B*, chr30:32,662,210).

In the neurocranium centroid size dataset, a marker at the CFA1 *SMOC2* locus is close to achieving genome-wide significance whilst a suggestive peak is emerging

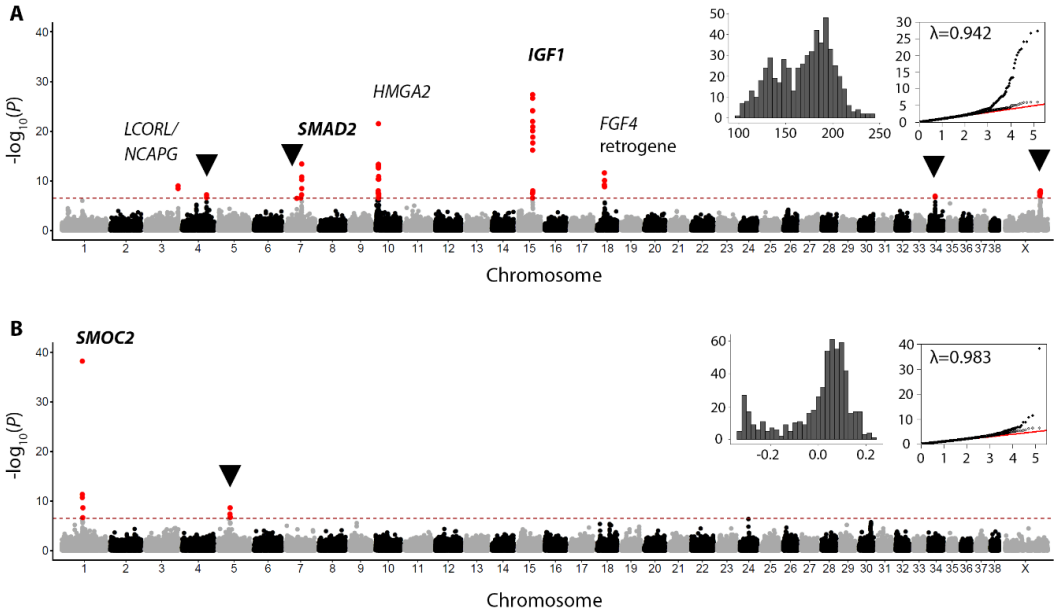


Figure 4.4 – Study Cohort 2 GWAS. Manhattan plots for the (A) neurocranium centroid size and (B) viscerocranium PC1 score for the second study cohort. Markers surpassing the threshold (dashed line) for genome-wide significance as calculate by a Bonferroni correction are coloured red. Positional gene candidates are given above each QTL with bold text denoting intragenic index markers. Additional QTL identified in this study compared to the first study population are marker by black arrows. Insets: Distribution plots displaying the frequency (y-axis) of the phenotype scores (x-axis).QQ-plots for the predicted (x-axis) and observed (y-axis) -log$_{10}$[p-values] for association tests. P-values plotted using the entire marker dataset (black circles) and pruned dataset following the removal of significant markers and markers in high linkage disequilibrium (r2>0.2). Abline (y=x) plotted in red. Genomic inflation value (λ) shows minimal inflation across analyses.

at the start of CFA1 (chr1:16,255,217) (Figure 4.4A). Interestingly, the melanocortin 4 receptor (*MC4R*) is in close proximity to this region which is known to be a key regulator of body weight (Huszar *et al.* 1997; Ollmann *et al.* 1997). Furthermore, the suggestive peak emerging on CFA4 (chr4:39,159,873) resides within the *STC2* gene.

4.3.3.2 Defining Critical Intervals

Viscerocranium Shape

The second, larger study cohort provides an opportunity to revisit and define new critical intervals for QTL associated with changes both viscerocranial shape and neurocranium centroid sizes. For the viscerocranial dataset, the index marker on CFA1 continues to tag the large selective sweep in the region with 40 markers being in high linkage disequilibrium (LD) ($r^2 > 0.2$) (Figure 4.5A). Despite this, the inclusion

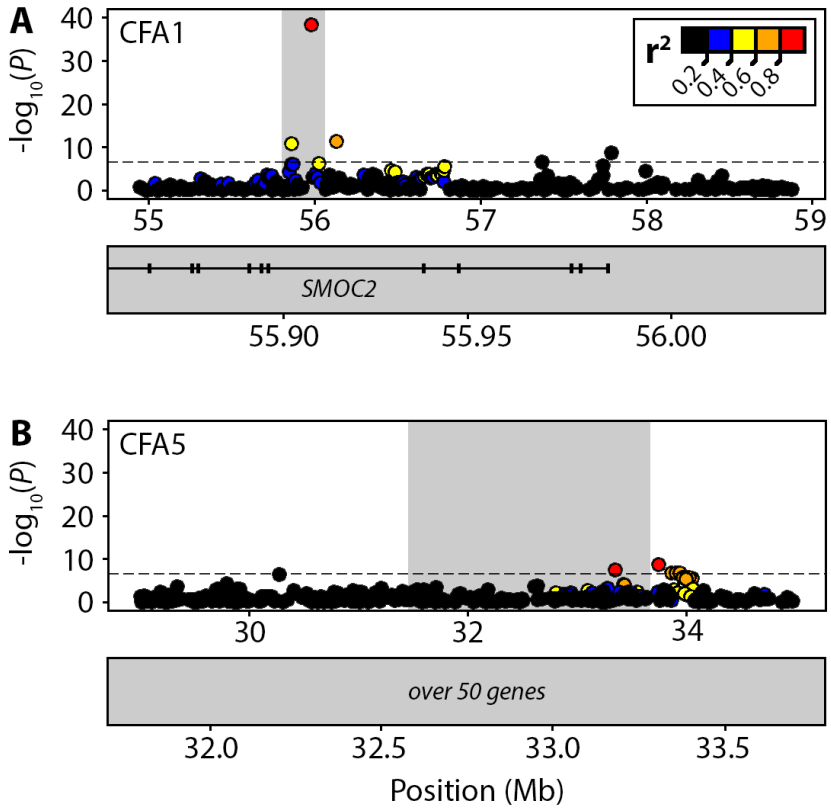


Figure 4.5 – Defining Critical Intervals for Brachycephaly – Zoomed Manhattan plot for QTL on (A) CFA1 and (B) CFA5. Markers are coloured by the degree of linkage disequilibrium with the index marker. Haplotype analysis identified a critical interval in grey (top box) which with an overview of the annotated genes within this interval (box below).

of more brachycephalic breeds in the study cohort, particularly French Bulldogs which have introduced new, previously unseen haplotypes in the region permitted for the CFA1 critical interval to be refined further (Table 4.5). The interval now spans 175,640 bp and continues to encompass the *SMOC2* gene and the variants called within in Section 4.3.2.3. Interestingly, two additional markers surpassing genome-wide significance were identified ~1 Mb downstream of the index marker (chr1:57,371,701 and chr1:57,790,010). These markers are not in LD with the index marker and it is unclear whether these they are tagging a second QTL on CFA1 or are an extension of that already defined. The vestigial like family member 2 (*VGLL2*), proto-oncogene receptor tyrosine kinase 1 (*ROS1*) and discoidin, CUB and LCCCL domain containing protein 1 (*DCBLD1*) genes are annotated here.

On CFA5, a total of 49 markers are in high LD with the index marker and span up to a ~1 Mb upstream of its position (Figure 4.5B). This extensive LD is reflected in the haplotypes in that whilst the index marker is positioned within the *NTRN1* gene, the haplotype that encompasses it is vast and cannot be refined further from ~2 Mb interval (Table 4.5). These large haplotypes are enriched in breeds such as the French Bulldog, French Mastiff and Boxer but are completely absent in the Pug. There are >50 genes annotated in this region.

CFA	Critical Interval	Size (bp)	Gene Candidate
1	55,862,036 – 56,037,676	175,640	*SMOC2*
5	31,714,339 – 33,788,841	2,074,502	*NTRN1*
3	91,097,987 – 91,327,127	229,140	*LCORL/NCAPG*
4	66,984,116 – 67,148,328	164,212	*GHR*
7	30,178,588 – 30,308,120	129,532	*TBX19*
7	43,515,166 – 43,948,370	433,204	*SMAD2*
10	8,270,435 – 8,397,696	127,261	*HMGA2*
15	41,188,905 – 41,232,547	43,642	*IGF1*
34	18,475,392 – 18,737,870	262,478	IGF2BP2

Table 4.5 – Critical Interval Summary. A list of critical intervals and their sizes defined at each of the morphological QTL identified in the second study cohort. Positional gene candidates are listed as the only genes annotated within the critical interval. In instances where multiple genes are annotated within the critical interval, the positional gene candidate is that closest to the index marker.

Neurocranium Centroid Size

Two markers on CFA3 surpassed genome-wide significance for the neurocranium centroid size which reside upstream of the *LCORL* gene. These two markers represent the extent of LD within the region and are enriched in large breeds such as the French Mastiff, Bernese Mountain Dog, Great Dane and Akita (Figure 4.6A). The *LCORL*, *NCAPG* and *ANAPC13* genes are annotated in their entirety within the 436,887 bp critical interval (Table 4.5, Figure 4.6A).

A further two significant markers are positioned on CFA4 which show minimal LD with markers in their immediate vicinity (< 400 Kb) (Figure 4.6B). The relatively small critical interval (164,212 bp) encompasses a haplotype that is enriched in many small breed dogs including several of the terriers (Cairn, Maltese, Border) as well as breeds such as the Dachshund and Bichon Frise (Table 4.5). Approximately 125 Kb of the 5' end of the *GHR* gene is situated within this haplotype in addition to the 3' end of the coiled-coil domain containing protein 152 (*CCDC152*).

Three markers are in modest LD with the single index marker at the first QTL on CFA7 (Figure 4.6C). A small haplotype at this locus, 129,532 bp is size is enriched for toy breed dogs such as the Chihuahua, Papillion and Pekingese however it also exists in unexpected breeds such as the Rhodesian Ridgeback, German Shepherd Dog and the Husky (Table 4.5). It is unclear why the haplotype would exist across these breeds. Since the critical interval is small, the haplotype in these large breeds may represent an ancestral version of the haplotype that hasn't been broken or lost. This critical interval encompasses the *TBX19* gene as well as a vesicle transport protein, *SFT2D2*.

At the second QTL on CFA7, markers show LD over a broad range, approximately 1 Mb in length (Figure 4.6D). At the centre of this selective sweep is a 433,204 bp haplotype with an index marker residing in the *SMAD2* gene (Table 4.5). This haplotype is present in a range of small size dog breeds such as the Yorkshire Terrier, Pug, Chihuahua and Lhasa Apso.

Markers spanning up to 1.5 Mb on CFA10 display high LD with the index marker within the QTL (Figure 4.6E). Interestingly, the 127,261 bp critical interval does not encompass the index marker like all other defined critical intervals associated with morphological variation in this study (Table 4.5). The reason for this is unclear. Nonetheless, this haplotype is enriched in small dog breeds such as the Border terrier,

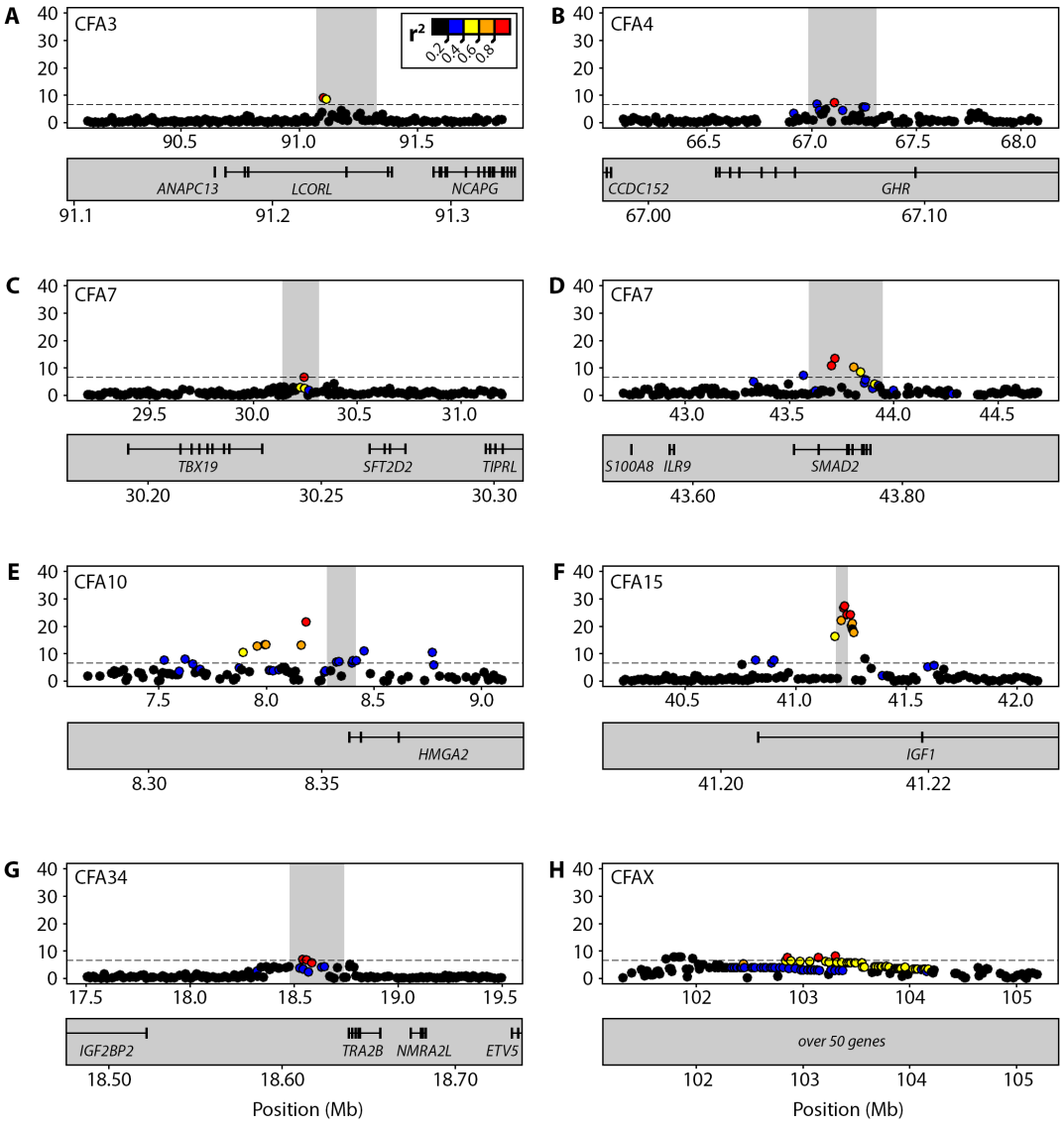


Figure 4.6 – Defining Critical Intervals for Size – Zoomed Manhattan plot for QTL on (A) CFA3, (B) CFA4, (C,D) CFA7, (E) CFA10, (F) CFA15, (G) CFA34 and (H) CFAX for neurocranium centroid size. Markers are coloured by the degree of linkage disequilibrium with the index marker. Haplotype analysis identified a critical interval in grey (top box) which with an overview of the annotated genes within this interval (box below).

Chihuahua, Jack Russel Terrier, and Yorkshire Terrier. The 5' end of the *HMGA2* gene and ~80 Kb of its upstream sequence is within this interval.

The QTL on CFA15 is largely comprised of eleven densely positioned markers in high LD all of which surpassing genome-wide significance around the index at chr15:41,221,438 (Figure 4.6F). A further three markers were in LD with the index and also surpassed genome-wide significance albeit ~500 Kb upstream. The critical interval at this locus is the smallest defined in this study, spanning just 43,642 Kb (Table 4.5). Approximately 29 Kb of the *IGF1* gene is captured by this haplotype which is enriched amongst many small dog breeds.

Two markers on CFA34 were significantly associated with neurocranium centroid size. Both these two markers were in very high LD with one another as well as a seven markers within the vicinity (Figure 4.6G). The haplotype is predominantly enriched in a small number of terrier breeds such as the West Highland White Terrier, Cairn Terrier and Maltese Terrier. The 262,478 bp critical interval encompasses the 5' end of the *IGF2BP2* gene in addition to its neighbouring gene, transformer 2 beta homolog protein (*TRA2B*), the latter of which has been linked with both body mass index (Table 4.5) (Pihlajamäki *et al.* 2011; Kaminska *et al.* 2016).

Finally, ten markers surpass genome-wide significance for neurocranium centroid size on CFAX. These markers span over 1.5 Mb and the index marker alone is in high LD with 78 other markers on the X chromosome (Figure 4.6H). It is unclear whether these markers are tagging a single, very large QTL or is in fact made up of at least two smaller QTL. As a result, it isn't possible to define a critical interval on the X chromosome and so many compelling positional candidate genes exist. The extensive haplotype sharing is explained by the reduced number of recombination events that can occur on the X chromosome due to YX haplodiploidy in males, which is then exaggerated by the frequent use of popular sires.

4.4 Discussion

Common dog traits such as brachycephaly are believed to predate the establishment of the regimented breed boundaries and are thought to be the result of identity-by-descent genetics. Due to the complex nature of these traits, the occasional 'outlier' within breeds which are not fixed for the causative variant have undermined previous attempts to resolve the QTL. Furthermore, extensive linkage disequilibrium within breeds has prevented the separation of co-inherited variants. To avoid this, for the first time this study has utilised individual-level phenotypes including those of mixed-breed dogs to distil our understanding of morphological QTL.

4.4.1 Face Shape

The selective sweep on CFA1 associated with canine brachycephaly has been recognised since the early days of dog GWAS however the resolution of the locus and its causal variant have remained elusive until now (Jones *et al.* 2008; Bannasch *et al.* 2010; Boyko *et al.* 2010; Schoenebeck *et al.* 2012). Here, a ~187 Kb haplotype encompassing the *SMOC2* gene was observed to be enriched in brachycephalic dogs. *SMOC2* is a compelling positional candidate gene for canine brachycephaly. It belongs to a family of matricellular proteins – the BM-40 (or osteonectin) family which were first identified to be expressed in bone (Whitson 1981; Vannahme *et al.* 2003). Since then, *SMOC2* has been shown to be expressed across a variety of tissues where it is thought to facilitate the remodelling of extracellular matrix by interacting with various growth signals (Göhring *et al.* 1998; Kupprion *et al.* 1998; Brekken and Sage 2000; Zhang *et al.* 2008; Maier *et al.* 2008; Feng *et al.* 2009). Loss of *smoc2* in zebrafish causes midface hypoplasia whilst different missense mutations within the gene have been implicated with severe dental abnormalities in humans (Bloch-Zupan *et al.* 2011; Alfawaz *et al.* 2013; Melvin *et al.* 2013). Here, five intronic variants including four SNPs and one structural variant (LINE-1) within *SMOC2* were associated with brachycephaly. Interestingly, following the increase in size of the study population, the critical interval surrounding at the *SMOC2* locus was further

refined from ~187 Kb to ~175 Kb highlighting the power and importance of finding rarer haplotypes which have undergone meiotic recombination. Despite this, the same five variants remained within this new interval. The functional effects of these variants on the transcriptomic profiles of brachycephalic dogs as well as modelling of their phenotypic effects sizes are investigated in Chapter 5.

Beyond CFA1, a second signal reaching genome-wide significance was detected on CFA5 with an index marker residing in the *NTRN1* gene (chr5:33,750,479). Two previous independent studies similarly identified a CFA5 QTL within this region however it was situated in an intergenic region 4.4 Mb upstream (Boyko *et al.* 2010; Schoenebeck *et al.* 2012). Due to the very large ~2.1 Mb critical interval defined in this study, it is possible that both of these markers are tagging the same selective sweep. Whilst the critical interval could not be more accurately defined, *NTRN1* itself is a compelling positional gene candidate. *NTRN1* is a diffusible signalling protein found in the extracellular matrix which is best characterised as a potent axonal guidance molecule (Cirulli and Yebra 2007; Masuda *et al.* 2009). Recent studies have suggested a more varied function for *NTRN1* including roles in the morphogenesis of various organs such as the ear (Salminen *et al.* 2000), pancreas (Hebrok and Reichardt 2004) and lungs (Liu *et al.* 2004) where it acts through regulating cell adhesion, migration and survival (Cirulli and Yebra 2007). Specifically with regards to craniofacial development, *NTRN1* has been shown to be strongly expressed in the palatal shelves (Leslie *et al.* 2015) with several associations between the gene and risks of developing non-syndromic cleft lip and/or cleft palate (Beaty *et al.* 2010; Sun *et al.* 2015; Velázquez-Aragón *et al.* 2016). Interestingly, albeit in murine cochlear hair cells, the *NTRN1* gene is expressed as a direct result of *IGF1* signalling which could provide a fascinating interaction between pathways predicted to regulate both craniofacial shape development and overall body size in dogs (Hayashi *et al.* 2013; Yamahara *et al.* 2017).

Furthermore, a collection of compelling positional gene candidates underlie suggestive associations with craniofacial shape variation including *HDAC5*, *GLI3*, *ASIP*

and *CORO2B*. Most strikingly in this list is *GLI3* – a gene known to cause Greig syndrome, a form of craniosynostosis characterised by a prominent forehead, wide-spaced eyes and a broad nasal bridge in humans (see Section 1.3.1) (Wild *et al.* 1997; Kalff-Suske *et al.* 1999). Meanwhile, *HDAC5* belongs to a class of histone deacetylase enzymes which have been shown to limit osteoblast differentiation through the inhibition of *Runx2* signalling (Kang *et al.* 2005; Schroeder and Westendorf 2005) – a pathway already known to regulate bone maturation and craniofacial development across mammals (see Chapter 1) (Yoshida *et al.* 2004; Fondon and Garner 2004; Abzhanov *et al.* 2007; Takarada *et al.* 2015). The *ASIP* gene encodes the Agouti signalling protein which regulates coat pigmentation across domestic species including the cow (Girardot *et al.* 2005), pig (Drögemüller *et al.* 2006), horse (Rieder *et al.* 2001) and dog (Berryere *et al.* 2005; Dreger and Schmutz 2011). Up to six distinct coat variations have been attributed to alleles at this locus and mice have been shown to have transcripts with four alternate first exons (Vrieling *et al.* 1994; Dreger and Schmutz 2011). Beyond pigmentation, a dominant Agouti allele in mice results in the ubiquitous expression of the protein and the onset of obesity (Huszar *et al.* 1997; Ollmann *et al.* 1997). The same mutant mice also have a marked increase in linear growth, giving them a long anteroposterior axis whilst also reducing the size and shape of the mandible (Suto 2008; 2009). These morphological observations coupled with the wide variation in pigmentation highlight the complex nature of the Agouti locus. A role for the neuronal tissue-enriched actin cytoskeleton regulator, *CORO2B* in craniofacial development is less clear due to the limited studies (Nakamura *et al.* 1999; Rogg *et al.* 2017).

This data speculatively supports the role of *GLI3*, *HDAC5* and *ASIP* in canine craniofacial morphological variation. Unfortunately, due to the apparent rarity of these alleles in this data (or limited phenotypic size effect), it makes it difficult to speculate further on the functional roles these genes may have in the dog. In comparative human studies, a common approach to validate the role of QTL in a given trait is to split the study population into discovery and replication cohorts (Liu *et al.* 2012; Adhikari *et al.* 2016). Such an approach can aid in the statistical

interrogation of QTL but have required considerable population sizes (>2,000). Like the many other canine studies, this thesis has not adopted a discovery and replication experimental design simply due to the limited number of participants (~500) (Jones *et al.* 2008; Bannasch *et al.* 2010; Boyko *et al.* 2010; Schoenebeck *et al.* 2012). If the study population for this experiment continues to grow, it may warrant the inclusion of a replication cohort in subsequent tests which may give more power to interrogate each of these QTL.

4.4.2 Body Size (Neurocranium Centroid Size)

By necessity, this study was unable to assess changes in canine craniofacial shape without first addressing the confounding influences of allometry. Here, the neurocranium centroid size was used a proxy to skeletal size which also provided an opportunity to explore the genetic contributions to the neurocranium itself. Following its application as a GWAS phenotypic input, a combined total of eleven significant and suggestive QTL were identified which have almost all been previously linked with body size variation within the dog and across species. This includes *IGF1* (Boyko *et al.* 2010; Vaysse *et al.* 2011; Rimbault *et al.* 2013; Hayward *et al.* 2016), *HMGA2* (Zhang *et al.* 2008; Boyko *et al.* 2010; Song *et al.* 2010; Rimbault *et al.* 2013; Fusco *et al.* 2015; Hayward *et al.* 2016), *SMAD2* (Boyko *et al.* 2010; Rimbault *et al.* 2013; Hayward *et al.* 2016), *GHR* (Ayling *et al.* 1997; Rimbault *et al.* 2013; Cui *et al.* 2015; Hayward *et al.* 2016), *STC2* (Gagliardi *et al.* 2004; Boyko *et al.* 2010; Rimbault *et al.* 2013) and the *LCORL*/*NCAPG* locus (Vaysse *et al.* 2011; Liu *et al.* 2013; Tetens *et al.* 2013; Sahana *et al.* 2015; Hayward *et al.* 2016). In the dog, variants tagging each of these loci have been reported with some interesting propositions such as a SINE in second intron of *IGF1* and a variant in the 5' UTR of *HMGA2*. However, the critical intervals surrounding these loci have never been explicitly defined in the literature despite profound efforts to fine map them (Rimbault *et al.* 2013). An exception applies to the *IGF1* locus which was refined to a 20-marker haplotype spanning ~65 Kb at chr15:41,205,752-41,271,109 (CanFam3.1). Here, using individual-level haplotype this interval has been reduced to ~43 Kb.

Other positional gene candidates associated with size include the melanocortin receptor (*MC4R*) which is antagonised by the Agouti signalling peptide. Perturbations, within this pathway have already been shown to regulate obesity (see above) (Huszar *et al.* 1997; Ollmann *et al.* 1997). Meanwhile, *TBX19* is a transcription factor expressed in the pituitary where its knockdown causes a loss of adrenocorticotropic hormone (ACTH) which in turn can modulate body weight (Pulichino *et al.* 2003; Eriksson and Mignot 2009). As an interesting side note, alterations in the hypothalamus-pituitary-adrenal axis can also affect behaviours and have been hypothesised to at least partly be driving domestication across species (Fu *et al.* 2016; Hekman *et al.* 2018; Fallahshahroudi *et al.* 2018). In the case of *TBX19*, it is unlikely that behavioural traits have been captured in the neurocranium size. Finally, *IGF2BP2* (also abbreviated *IMP2*), which is involved in the post-transcriptional regulation of *IGF2* and is expressed throughout the adult mouse (Yisraeli 2004; Dai *et al.* 2011). Knockout *IGF2BP2* mice display an increased resistance to diet induced obesity and are much leaner than their wildtype littermates (Dai *et al.* 2015). However, focussing on the *IGF2BP2* locus highlights one of the fundamental problems in the field of canine genetics. A marker (chr34:18,559,537) positioned ~18 Kb from the CFA34 index marker reported here (Table 4.3) was previously associated with body weight by Hayward et al., (Hayward *et al.* 2016). Their study performed no haplotype analysis or attempts to refine a critical interval and yet the published results labelled the underlying causal gene as *IGF2BP2*. As highlighted here in Figure 4.6G, now that fine mapping of the locus has been performed, the critical interval encompasses another gene, *TRA2B* in its entirety. This gene is equally compelling to cause changes in body weight as reduced *TRA2B* expression and alternate splicing of its second exon are associated with changes in weight in humans (Pihlajamäki *et al.* 2011; Kaminska *et al.* 2016). The use of gene candidate names to identify QTL is understandable and simplifies the use of a coordinate-based system however without adequate fine mapping or the identification of causal variants, the labelling of loci with gene names in the context is misleading and should be avoided wherever possible.

Associations on the X chromosome have been linked to variation in canine body size previously however their resolution have proven extremely difficult due to the high homozygosity along the chromosome following the propagation of popular sires (Boyko *et al.* 2010; Vaysse *et al.* 2011; Hayward *et al.* 2016; Plassais *et al.* 2017). Two of these studies have highlighted the immunoglobulin superfamily member 1 (*IGSF1*) as a compelling candidate to be associated with large dog breeds. Other noteworthy findings include the insulin receptor substrate protein 4 (*IRS4*) and acyl-CoA synthetase long-chain family member 4 (*ACSL4*) genes to be associated with large dogs and heavily muscled dogs respectively. These associations have putatively causal variants listed with them however their implications on gene function such as transcriptomics were not analysed. The association in this current study has an index marker within the *MBNL3* gene which is less than 1 Mb away from any of the *IGSF1* and is thought to antagonise myogenesis in mice (Poulos *et al.* 2013). It is possible that whilst both of these positional gene candidates are compelling, the markers from both studies may in fact be tagging the same selective sweep in which case further assessment at this locus is needed. Interestingly, before these studies, a broad interval spanning both the *MBLN3* and *IGSF1* genes has also been linked to driving sociability and aggressiveness in dogs (Vaysse *et al.* 2011; Zapata *et al.* 2016). It remains unclear whether the same positional gene candidates could drive body size and behavioural traits variation or if multiple causal genes are co-segregating. Ultimately, both reduced size and aggression have been characteristics selected for throughout domestication.

It is intriguing that many of the known functions of these genes are to drive alterations in body weight or muscling whilst the phenotype used here to identify these loci has its origins in the size of the neurocranium – a skeletal structure. Changes in skeletal structure may alter body weight however the inverse is not necessarily true. A possible hypothesis is that these positional gene candidates have roles beyond soft tissue (muscling) and metabolic (weight) effects and in fact also drive changes in skeletal structure too. An alternate hypothesis is that there is a

cryptic covariation between dogs with increased mass and those with larger neurocranium. Such a relationship is currently unclear.

Finally and unexpectedly, the *FGF4* retrogene locus on CFA18 was associated with neurocranium centroid size. Previously, the locus was first identified to control asymmetric chondrodysplasia across short-legged breeds such as the Basset hound (Parker *et al.* 2009). The locus was subsequently associated with variation in canine body weight but as discussed above, this could be explained by shorter bones and therefore lower leg mass in dogs (Hayward *et al.* 2016). This data suggests for the first time that the *FGF4* retrogene may also be involved in the formation and therefore variation of bony structures such as the skull.

Collectively, these data must, like all genomic association studies, be taken at face value until causal variants can be identified and their functional effects fully characterised. Here, compelling positional candidate genes based upon their known biological functions have been discussed, however, until further downstream experiments are carried out these genes are purely speculative. Furthermore, whilst clear associations have been drawn and their underlying critical intervals defined, in some instances it remains unclear precisely what morphological trait is driving them. For example, are all the viscerocranium PC1 associations being driven by the shortening of the face or is there some cryptic cosegregation with another trait? Could viscerocranium PC1 also be capturing traits such as klinorhynchy which is a common feature of brachycephalic dogs? The Agouti locus emerged as a suggestive peak – is this determining head shape or could the brachycephalic dog breeds in this study have an overrepresentation of a specific coat colour variant? The ambiguity in phenotypes can also extend to the bony neurocranium centroid which has identified loci previously known to drive soft tissue traits such as muscling and body weight.

In order to confirm that the CFA1 locus associated with canine brachycephaly is indeed dictating facial length, Chapter 5 will investigate the role of the five remaining intronic variants in the *SMOC2* gene and their predicted morphological effects

Chapter 5: Characterisation of *SMOC2* on Craniofacial Variation

5.1 Introduction

In the previous chapter, a critical interval for canine brachycephaly was defined on chromosome 1 which encompassed the *SMOC2* gene. Subsequent variant calling and filtering identified five intronic variants within the gene which represent putatively causal variants for canine craniofacial variation. This variant list included four SNPs and the insertion of a truncated long interspersed nuclear elements class 1 (LINE-1).

LINE-1s are a highly abundant form of retrotransposon thought to make up as much as 19% of the dog genome (Lindblad-Toh *et al.* 2005). Full length LINE-1 retrotransposons can be up to ~6 Kb in length and contain two open reading frames (ORF) (Beck *et al.* 2011). ORF1 encodes a nucleic acid binding protein which has been shown to have a nucleic acid chaperone activity (Callahan *et al.* 2011). Meanwhile, ORF2 encodes a protein with endonuclease and reverse transcription activities which enables the elements to mobilise in a "cut and paste" fashion with preferential targeting of TTAAAA hexanucleotide insertion sites (Feng *et al.* 1996; Jurka 1997; Hohjoh and Singer 1997; Beck *et al.* 2011). Despite the high abundance of LINE-1 elements in mammalian genomes, very few are thought to be active, largely due to 5' truncations causing the loss of the ORFs during "abortive" retrotransposition or the introduction of inactivating mutations (Han *et al.* 2002; Gilbert *et al.* 2005; Kopera *et al.* 2016). In the average human it is thought that there may be between 80 and 100 full-length LINE-1s (Brouha *et al.* 2003) with approximately half which display highly active of "hot" retrotransposition capabilities (Brouha *et al.* 2003). Interestingly, there is a higher number of more recent LINE-1 activity in dogs compared to humans, which may be a by-product of the distinct canine domestication process and begin to explain the rapid diversification within the species (Wang and Kirkness 2005).

The mobilisation and reintegration of these large genetic elements around a genome can unsurprisingly have detrimental effects. Many LINE-1 elements are situated either in introns or intergenic regions but can have significant impacts beyond this (Szak *et al.* 2002). As a result of the transposition, LINE-1s can cause large deletions, genomic inversions, intrachromosomal duplications, interchromosomal translocations and introduction of additional non-LINE-1 sequence such as SINEs and retrogenes (Symer *et al.* 2002; Gilbert *et al.* 2005). If these genomic alteration occurs in proximity to genes, the function of those genes could be compromised. For example in humans, the insertion of a LINE-1 into an exon of the factor VIII gene results in a complete loss of function of the protein and causes haemophilia whilst the insertion of a different LINE-1 into the dystrophin gene causes exon skipping and the onset of Duchenne muscular dystrophy (Kazazian *et al.* 1988; Narita *et al.* 1993). The introduction of LINE-1 elements into introns can also modify transcripts either by offering alternative splice sites which has been shown to cause retinitis pigmentosa 2 and granulomatous disease (Schwahn *et al.* 1998; Meischl *et al.* 2000), or alternatively causing premature transcript truncation by introducing a polyadenylated tail (Perepelitsa-Belancio and Deininger 2003). Intronic LINE-1 elements have also been shown to cause a reduced expression of the endogenous gene, possibly by causing transcriptional elongation and the increasing the likelihood

Disease	Gene	Mechanism	Reference
Retinopathy	*MERTK*	Unknown	(Everson *et al.* 2017)
Cleft Palate	*DLX6*	Premature transcript truncation	(Wolf *et al.* 2014)
Duchenne-like Muscular Dystrophy	Dystrophin	Premature transcript truncation	(Smith *et al.* 2010)
Haemophilia B	Factor IX	Gene downregulation through unknown mechanism	(Brooks *et al.* 2004)
Canine Transmissible Venereal Tumour	c-myc	Unknown	(Choi *et al.* 1999)

Table 5.1 – LINE-1 Associated Diseases in the Dog. A list of diseases attributed to a LINE-1 insertion within or adjacent to positional candidate genes. When known, the mechanism through which the LINE-1 is believed to act upon its associated gene is listed.

of the RNA polymerase disengaging from the DNA (Han *et al.* 2004). Alternatively, any of the structural genomic modifications could affect upstream regulatory elements and modify gene expression.

Comparatively, a diverse range of dog diseases have been linked to the insertion of LINE-1 elements within or adjacent to genes (Table 5.1). The mechanism through which the LINE-1 insertions act upon their associated gene are less well characterised but nonetheless present compelling cases for causality of their respective diseases.

In this chapter, the objective is to test the hypothesis that *SMOC2* is the most compelling gene candidate for canine brachycephaly and that the insertion of the LINE-1 within the gene is the causative variant.

5.2 Materials and Methods

5.2.1 Genomic DNA Extraction from Gonadal Tissue (stored in RNAlater®)

In collaboration with Friendship Animal Hospital (Washington, DC USA), Dr. Schoenebeck collected gonad tissue from healthy dogs undergoing neutering surgery. Samples were donated with owner consent. Gonadal tissue was preserved in RNAlater® and stored at -80 °C. In order to permit a targeted mRNA extraction from testis, gDNA was first extracted and genotyped for the *SMOC2* LINE-1 insertion as described previously. The protocol for extracting gDNA from RNAlater®-stored tissue was adapted from ThermoFisher Scientific (http://www.thermofisher.com/uk/en/home/references/protocols/nucleic-acid-purification-and-analysis/rna-protocol/genomic-dna-preparation-from-rnalater-preserved-tissues.html). A digestion buffer was prepared containing 1.8 mL 1 M Tris (pH 8.0), 4.0 mL 0.5 M EDTA, 0.5 mL SDS (20%) and 13.7 mL deionised water. On lint-free paper, a testes samples approximately 0.5 cm^3 in size were cut using a scalpel and any remaining RNAlater was removed. The testes samples were minced well with a scalpel blade and placed in 2 mL microfuge tubes with 1.5 mL digestion buffer and 37.5 µL 20 mg/mL Proteinase K. The samples were vortexed for 2 seconds before

incubating on a heat block at 55 °C for 4 hours. Once digestion was complete, samples were divided into 750 μL volumes before adding 750 μL phenol chloroform (Thermo Fisher, 15593-031) and inverting rapidly for 2 minutes at room temperature. Samples were centrifuged at 16,000 G for 10 minutes. The DNA-containing upper aqueous layer was transferred to a fresh microfuge tube and the phenol chloroform separation step was repeated twice more. After the third phenol chloroform step, the upper aqueous phase was transferred to a fresh 1.5 mL microfuge tube containing 50 μL 3M sodium acetate (pH 5.2) (Amresco, E498) and 500 μL 95% ethanol. Samples were mixed by inversion and incubated at room temperature overnight to allow full gDNA precipitation. Samples were centrifuged at 16,000 G for 5 minutes, supernatant removed and pellets washed with 500 mL 70% ethanol for 5 minutes. Samples were centrifuged again and pellets allowed to air dry once the ethanol was removed. gDNA was rehydrated in 200 μL Tris EDTA.

5.2.2 Total RNA Extraction from Frozen Testes in RNAlater™

Testes were selected for messenger RNA (mRNA) extraction due to the unattainability of appropriate embryonic-stage tissue. *SMOC2* was assumed to be expressed in the testes based on evidence in other species (Pazin et al. 2009; Uhlén *et al.* 2015). In total, nine dogs were selected for RNA extraction representing three individuals for each of the *SMOC2* LINE-1 genotypes – homozygous ancestral (1 Italian Greyhound, 1 Whippet, 1 Yorkshire Terrier), 3 heterozygotes (1 Papillion, 2 Cavalier King Charles spaniels) and 3 homozygous derived (1 Bulldog, 1 French Bulldog, 1 Pug).

Prior to RNA extraction, all equipment was cleaned using RNaseZap™ (Ambion, AM9780). For each RNA preparation, each sample was weighed whilst frozen to ensure it did not exceed 100 mg. Any remaining RNAlater™ was removed using lint-free paper prior to adding the sample to 1 mL chilled TRIzol™ (Life Technologies, C-15596026EB) in a Matrix D lysis tube containing 1.4 mm ceramic spheres (Fisher Scientific, MBR-247-110Y) on wet ice. Lysis tubes were secured in the FastPrep™-24 device and run at a speed of 4 m/s for 20 seconds. Homogenised samples were stored on ice before being incubated at room temperature for 5 minutes. Samples were

shaken vigorously for 15 seconds with 200 μL BCP (Sigma, B9673) and incubated for a further 3 minutes at room temperature. Samples were centrifuged at 12,000 G for 15 minutes at 4 °C and the upper RNA-containing aqueous phase was transferred to a fresh microfuge tube. Total RNA samples were cleaned using the RNeasy Lipid Tissue Mini Kit (Qiagen, C-74104EB) following the manufacturer's instructions. Briefly, one volume of 70% ethanol was added to the RNA samples and vortexed for 2 seconds. A maximum of 700 μL sample was added to a RNeasy Mini spin column in a 2 mL collection tube and spun at 10,000 G for 15 seconds at room temperature. The effluent was discarded and any remaining sample was re-run using the same column. Buffer RW1 (700 μL) was added to the column and centrifuged at 10,000 G for 15 seconds – the effluent was discarded. Buffer RPE (500 μL) was added to the column, spun for 15 seconds and the effluent was discarded. A further 500 μL of RPE was added to the column, spun for 2 minutes and the effluent was discarded. The column was spun for a further 2 minutes to remove residual buffer. The column was placed in a 1.5 mL collection tube and 30 μL water was added. The column was left at room temperature for 30 seconds before the centrifuging at 10,000 G for 1 minute to elute the RNA. The manufacturer's optional DNase step was not used.

5.2.3 Total RNA Quantification and Quality Assessment

Quantification was performed using the Qubit® RNA Broad Range Assay Kit (Life Technologies, Q10210). One 0.5 mL microfuge tube was labelled per sample and two for standards. Working solution was prepared in a plastic bijoux tube by diluting the Qubit® RNA BR Reagent 1:200 in Qubit® RNA BR buffer. The final volume of working solution accommodated for 200 μL per microfuge tube prepared previously. To prepare the standards, 10 μL of the appropriate Qubit® standard was added to 190 μL working solution in the labelled microfuge tubes and mixed by vortex for 2 seconds. To prepare the samples, 1 μL of each sample was added to 199 μL working solution in the appropriately labelled microfuge tubes and mixed by vortex for 2 seconds. Samples were incubated for 2 minutes at room temperature. All steps were

carried out with minimal exposure to light. The RNA Broad Range setting was selected on the Qubit® device, standards were read followed by samples.

RNA quality was quantified using the TapeStation 2200 (Agilent Technologies). RNA ScreenTape sample buffer (Agilent Technologies, 5067-5577) and RNA ScreenTape Ladder (Agilent Technologies, 5067-5578) were allowed to equilibrate to room temperature for 30 minutes before vortexing for use. The working ladder was prepared by mixing 5 μL RNA sample buffer with 1 μL RNA ladder. Total RNA samples were thawed on ice before 1 μL was added to 5 μL RNA sample buffer. Tubes were spun down and vortexed for 1 minute at 2000 rpm. Tubes were spun down and incubated at 72 °C for 3 minutes and then placed on ice for 2 minutes. Samples were spun down and loaded into the TapeStation instrument with fresh pipette tips and an RNA ScreenTape (5067-5576). An RNA integrity number equivalent (RINe) of ~8.0 and above is considered to represent RNA with minimal degradation.

5.2.4 RNA Sequencing and Analysis

In total, nine RNA samples were selected for sequencing and included three representatives from the three *SMOC2* LINE-1 genotypes – homozygous ancestral (1 Italian Greyhound, 1 Whippet, 1 Yorkshire Terrier), 3 heterozygotes (1 Papillion, 2 Cavalier King Charles Spaniels) and 3 homozygous derived (1 Bulldog, 1 French Bulldog, 1 Pug). All RNA samples exceeded RINe values of 0.8 and were submitted to Edinburgh Genomics, UK for sequencing. Briefly, TruSeq standard RNA libraries were prepared for each of the nine total RNA samples following the manufactures guidelines. Barcoded libraries were sequenced on three lanes of an Illumina HiSeq 4000, producing 150 bp paired-end reads.

The Splice Transcripts Alignment to a Reference (STAR) software (v2.5.1b) was used to align FASTQ files containing the RNA short-read data to an indexed CanFam3.1 reference sequence (CanFam3.1, ENSEMBL release-85) using default parameters (Dobin *et al.* 2012). Annotated junctions were downloaded from ENSEMBL. Read alignment was performed in two passes as recommended by the

manufacturer. Picard tools (v2.18.5) was used to add read groups, merge .bam files by sample and mark duplicate reads (http://broadinstitute.github.io/picard). FeatureCounts, an analysis tool of the RSubread package (RSubread v1.22.3 installed on R), was used to quantify mapped reads to genes (Liao *et al.* 2013). Differential expression analysis was conducted at the gene level using the R package DESeq2 (v1.12.4) by comparing three homozygous *SMOC2* LINE-1 carriers compared to three non-carriers (Love *et al.* 2014). Detection of allelic imbalance in gene expression was made possible by three heterozygous dogs (2 Cavalier King Charles spaniels and 1 Papillion) which had a silent exonic single nucleotide variant (SNV) segregating with the LINE-1 insertion. RNAseq data was visualised using the Integrative Genomics Viewer (IGV) (Robinson *et al.* 2011).

5.2.5 Reverse Transcription

Complementary DNA (cDNA) was produced from 1 μg total RNA using the SuperScript III First-Strand SuperMix kit (Invitrogen). One μL 50 μM oligo(DT) primer was mixed with 1 μL annealing buffer and 1 μg total RNA and made up to 8 μL with DNase/RNase-free water. Samples were incubated in a thermocycler at 65 °C for 5 minutes in and then immediately placed on ice for a minimum of 1 minute. The samples were collected by briefly centrifuging before adding 10 μL 2X First-Strand Reaction Mix and 2 μL SuperScript™ III enzyme mix on ice. Samples were vortexed and collected by centrifuging before incubating at 50 °C for 50 minutes. The reaction was terminated at 85 °C for 5 minutes before chilling on ice.

5.2.6 Real-Time Quantitative Polymerase Chain Reaction (RT-qPCR)

5.2.6.1 Primer Design

In order to design primers for RT-qPCR the predicted transcript sequences were downloaded from the Ensembl genome browser (http://www.ensembl.org/index.html) for *SMOC2* (ENSCAFT00000001348.4), *DACT2* (ENSCAFT00000050141.2) and *THBS2* (ENSCAFT00000046608.2). Downloaded mRNA sequences were annotated by including square brackets, "[]" between each exon-

Target	Primer	Probe	Source
SMOC2 E2/3	(F)TGCTTATCGAGGAAATTGCAG	#68	(Marchant *et al.* 2017)
	(R)TGGGATGAACACCTGCTGTA		
SMOC2 E10/11	(F)CGCGCTCTCTACCGACAT	#73	(Marchant *et al.* 2017)
	(R)GGGGTCGGGTTCTGAGAG		
DACT2	(F)AGGACGTCGGCCTAAAAAC	#125	(Marchant *et al.* 2017)
	(R)CCACCATCACTCAGCTCGTA		
THBS2	(F)CTGGGCAAGTGCGACAAT	#19	(Marchant *et al.* 2017)
	(R)AAGACCATGGCGACCAGT		
MRPS7	(F)GGAGAAGAAGCAC	#89	(Maccoux *et al.* 2007)
	(R)GTGTGACAACT		

Table 5.2 – RT-qPCR primers. A list of exon-spanning primers designed to target specific genes. Primers were designed using the Roche Assay Centre which recommended the appropriate Roche Universal Probe to use.

exon boundary as indicated by colour changes in the Ensembl transcripts. Annotated sequences were uploaded to the Roche Universal Probe Library's Assay Design Centre (https://lifescience.roche.com/en_gb/brands/universal-probe-library.html#assay-design-center). The software does not have a canine-specific setting and so the "Other organism (paste sequence only)" option was selected with automatic selection of intron-spanning primer design being selected (Table 5.2). The primer sequences for the housekeeping gene, *MRPS7* were acquired from Maccoux et al., 2007 (Maccoux *et al.* 2007). All lyophilised primers were resuspended in deionised water to a final concentration of 100 µM. Probes from the Roche Universal Probe Library were recommended for each primer pair.

5.2.6.2 Primer Efficiency

All RT-qPCR was performed using the Roche LightCycler® 480 device with the LightCycler® 480 Multiwell Plate 96 following the manufacturer's instructions. For each reaction, 4.7 µL cDNA was added to 5 µL Roche Probe Master Mix, 0.1 µL 10 µM forward & reverse primer mix and 0.2 µL of the appropriate probe. Serial dilutions of prepared cDNA samples were used to optimise the protocol with technical triplicates being used. Table 5.3 lists the thermocycler conditions including a pre-incubation step for the un-purified cDNA.

Stage	Temperature (°C)	Time	Cycles
Pre-incubation	95	10 minutes	x1
Amplification	95	10 seconds	x45
	60	30 seconds	
	72	1 second	
Cooling	40	30 seconds	x1

Table 5.3 – LightCycler® Programme. Temperature and duration of stages in the standard LightCycler® programme for RT-qPCR.

Raw text files containing crossing point (Cp) values were exported from the Roche LightCycler® software and imported into Microsoft Excel for analysis. The slope between Cp and Log_{10} cDNA dilution was calculated using the "=SLOPE" function in Excel. Primer efficiency was then calculated using Equation 5.1.

Under optimum conditions, a primer with 100% efficiency would generate two new templates for every reaction cycle. Perfect efficiency is rare and so primers with efficiencies between 1.93 and 2.07 were considered suitable for transcript quantification.

$$Efficiency = 10^{\frac{-1}{Slope}}$$

Equation 5.1 – Amplification efficiency. The equation used to calculate the amplification efficiency of a primer pair using the slope of the curve between the crossing point (Cp) and the log_{10} cDNA dilution.

5.2.6.3 Relative Gene Expression

Only primers targeted at the housekeeping gene, *MRPS7* and the 5' and 3' targets of *SMOC2* had appropriate efficiencies for relative transcript quantification. RT-qPCR was performed using the master mix and thermocycler protocol as described previously (section 5.2.6.2). Biological and technical triplicates were used for cDNA dilutions of 1:100. Relative expression ratios (R) were calculated and corrected for primer efficiency using Equation 5.2 (Pfaffl, 2001).

$$R = \frac{Efficiency_{SMOC2}{}^{\Delta Cp\ SMOC2(Ancestral-Derived)}}{Efficiency_{MRPS7}{}^{\Delta Cp\ MRPS7(Ancestral-Derived)}}$$

Equation 5.2 – Efficiency-corrected Normalised Expression. Changes (Δ) in crossing point (Cp) scores for target (*SMOC2*) and reference (*MRPS7*). Δ Cp scores were calculated by subtracting the average for each gene in homozygous LINE-1 (Derived) from homozygous non-LINE1 carriers (ancestral).

5.2.7 Evolutionary Conservation

The Evolutionary Conserved Region Browser (https://ecrbrowser.dcode.org/) was used to assess the degree of sequence conservation across species. The Human genome (hg19) was used as the base genome with *SMOC2* gene coordinates chr6:168,841,831-169,068,674. Default settings were used to compare the Zebrafish, Chicken, Mouse, Cow, Dog and Rhesus Macaque sequences.

5.2.8 *In situ* Hybridisation

In situ hybridisations for mouse and chicken *Smoc2* were performed by Lynn McTeir, The Roslin Institute, The University of Edinburgh following the Nieto *et al.,* 1996 protocol (Nieto et al., 1996). Mouse (C57BL/6) and chicken (Isa Brown) embryos used for histology were surplus biomaterial harvested prior to this study. Mouse and chick work was performed in accordance to animal use guidelines of the Roslin Institute.

5.2.9 Mouse Skull Morphology Assessment

Lateral and dorsoventral radiographs of four C57BL/6NJ background controls and eight *Smoc2*$^{-/-}$ mice (Smoc2$^{tm1.1(KOMP)Vlcg}$ allele produced by The Jackson Laboratory, USA) aged 13 weeks were landmarked in ImageJ (v1.50g) using the PointPicker plugin (male = 5, female = 7) (Schneider *et al.* 2012). Nine lateral and fifteen dorsoventral landmarks of the murine skull were selected for their reproducibility across radiographs. The 2D landmark coordinates were exported from ImageJ and imported to MorphoJ. The lateral and dorsoventral landmarks were analysed separately but both used the same approach. A generalised Procrustes fit was used to appropriately superimpose, scale and orientate the landmarks. A

covariance matrix was calculated using the Procrustes distance matrix prior to principal component analysis. A two-tailed Student's t-test assessed the distribution of whole head PC1 scores for sex and *Smoc2* background. Additional phenotypic detail for these mice are available from the International Mouse Phenotyping Consortium (http://www.mousephenotype.org).

5.3 Results

5.3.1 Putatively Causal *SMOC2* Variants

The critical interval defined previously (Chapter 4) on CFA1 (55,850,299 - 56,037,676) for viscerocranium PC1 (brachycephaly) spans the *SMOC2* gene. Following SNP, small INDEL and large structural variant calling and hard filtering five putatively causal variants remained which included four SNPs and one large structural insertion (Table 5.4).

The remaining four SNPs are situated within introns 1, 7 and 12 of *SMOC2* whilst the structural variant resides in intron 8 (Figure 5.1A). The structural variant is a 1,531 bp insertion which is present in the brachycephalic Boxer reference genome. All four SNPs appear to be in complete linkage disequilibrium with the structural insertion. PhastCon scores downloaded from the UCSC genome browser indicate minimal (~0) conservation of each of the SNPs for which data was available for across the dog, mouse, rat and human genomes (Figure 5.1B). Suggestive that the SNPs are unlikely to be situated in a regulatory region, the candidacy for the SNPs to be putatively causal for canine brachycephaly is poorly supported from this data.

Software	Variant Type	Base Pairs Analysed	Pre-Filtering	Post-Filtering
GATK / SnpSift	SNPs / INDELS	187,377	3,674	4
Pindel	Structural Variants	187,377	162	1

Table 5.4 – Variant Filtering. Summary of variants called and filtered within the CFA1 critical interval.

The remaining 1,531 bp structural insertion is a 3' truncated fragment of a long interspersed nuclear element class 1 (LINE-1). Canine LINE-1 are large (>6,000 bp) retrotransposons known to be mutagenic in both dogs and humans (Miki *et al.* 1992; Lander *et al.* 2001). An intact LINE-1 is composed of two open reading frames (ORF), 5' and 3' untranslated regions (UTR) and a polyadenylated tail (poly(A)) (Figure 5.2A). During the complete retrotransposition of a LINE-1, target site duplications (TSD) are introduced and flank the insertion site. Here, the insertion associated with brachycephaly is a truncated, 1,506 bp fragment containing the 3' UTR and part of ORF2 in addition to a poly(A) tail which may have been the result of "abortive" retrotransposition (Gilbert *et al.* 2005). No target site duplications could be detected. The ancestral allele does not carry the LINE-1 insertion.

A multiplex touchdown PCR protocol was established to genotype all dogs used in the discovery phase GWAS using study cohort 1 for the *SMOC2* LINE-1 insertion. The LINE-1 insertion was found in 91.5% of brachycephalic dogs with viscerocranium

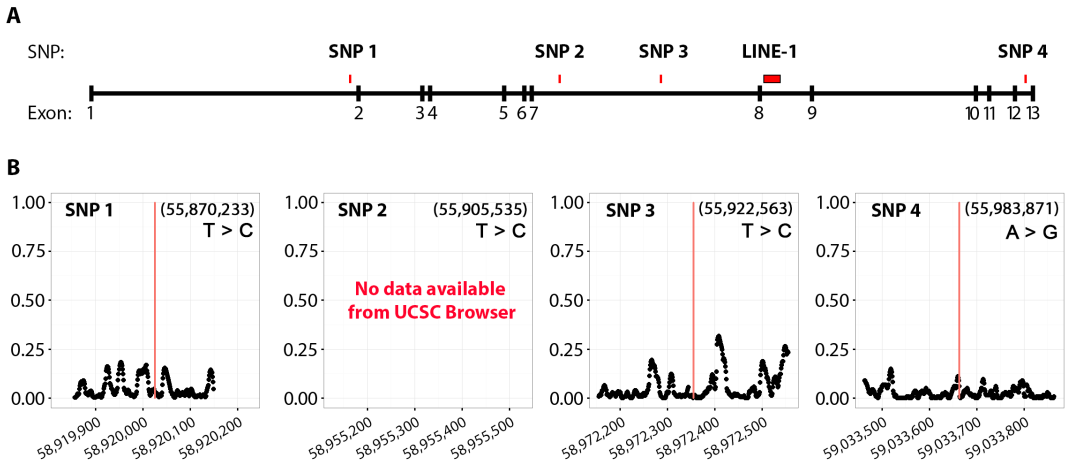


Figure 5.1 – Remaining Variants after Filtering. (A) All four remaining SNPs in the viscerocranium PC1 critical interval following filtering are located within the *SMOC2* gene. (B) Coordinates of each SNP (listed as CanFam3.1 in brackets) are converted to CanFam2.0 and used to plot PhastCon scores (y-axis) for 200 bp in both directions from the SNP (x-axis). PhastCon scores were downloaded from the UCSC Table Browser. No PhastCon scores were available for SNP2. From the available data, no SNPs reside in highly conserved positions across dog, human, mice and rat genomes. Alleles are written as Ancestral > Derived. Modified from: (Marchant *et al.* 2017).

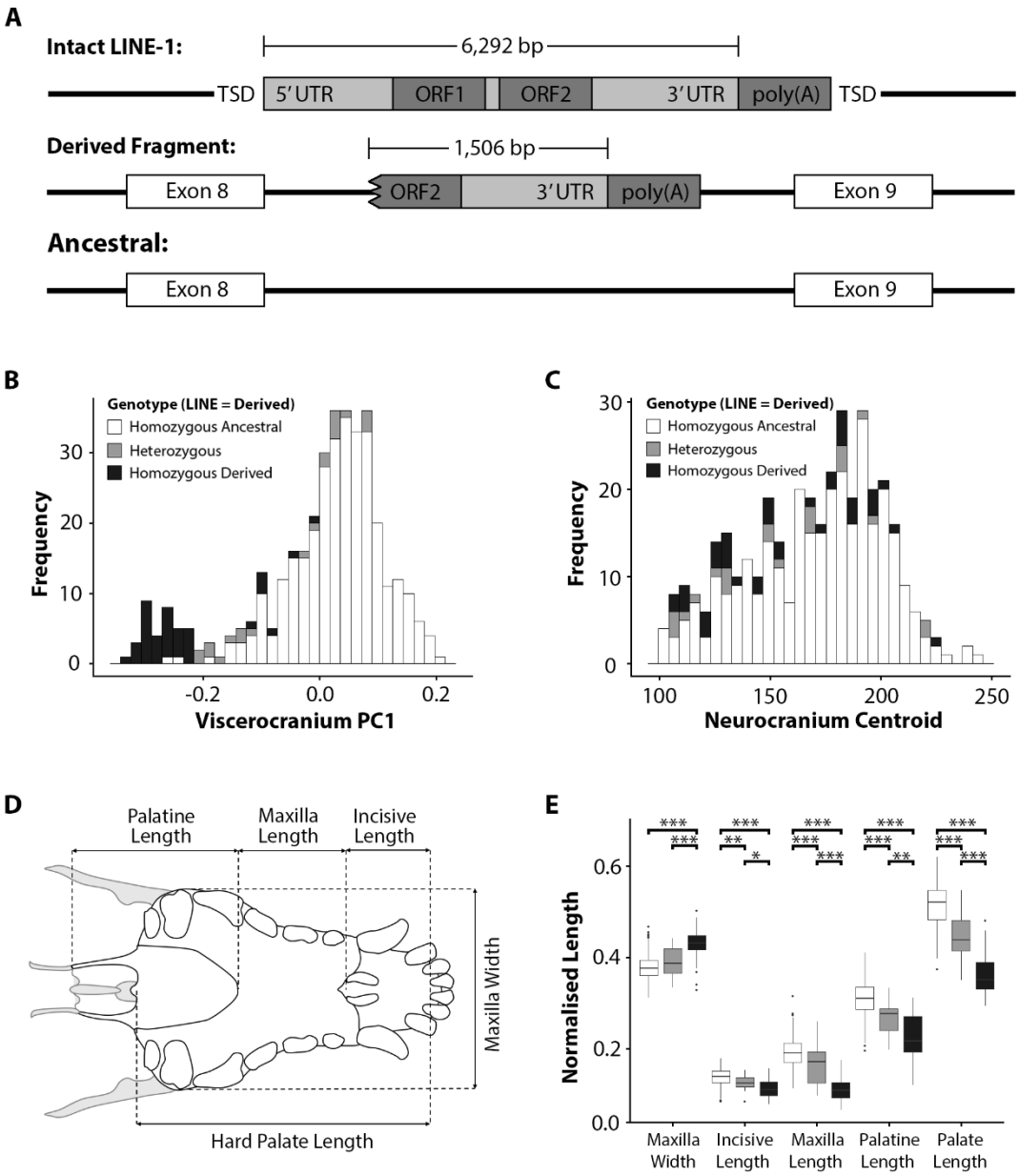


Figure 5.2 – *SMOC2* LINE-1 Distribution. (A) Schematic of a full-length canine LINE-1 element consisting of 5' and 3' untranslated region (UTR), open reading frames 1 and 2 (ORF1/ORF2) and a polyadenylated tail (poly(A)) flanked by target site duplications (TSD). The structural variant within SMOC2 is 1,506 bp in length (in addition to a poly(A) tail) and has a 99.1% match to the consensus sequence of a canine LINE-1. Distribution of the SMOC2 LINE-1 fragment for (B) viscerocranium PC1 and (C) neurocranium centroid size across all individuals. (D) Ventral-dorsal schematic view of a canine hard palate and its constituent bones. Schematic not drawn to scale (E) Length and width of the canine palate and constituent bones normalised by the neurocranium centroid size for homozygous ancestral (white), heterozygous (grey) and homozygous derived (black) individuals for the *SMOC2* LINE-1 insertion. Modified from: (Marchant *et al.* 2017).

PC1 scores < -0.2, whilst just 2.1% of mesocephalic to dolichocephalic (viscerocranium PC1 > -0.2) dogs (Figure 5.2B). Meanwhile, the LINE-1 insertion appears to have no correlation with the neurocranium centroid size (Figure 5.2C). To investigate whether the *SMOC2* LINE-1 genotype correlates with physical measurements of the skull, the length and width of the canine hard palate and its constituent bones were calculated using appropriate landmarks (Figure 5.2D). A dose-effect was observed across all linear measurements with two copies of the *SMOC2* LINE-1 correlating with shorter (along the body axis) and wider (perpendicular to the body axis) bones of the palate (Figure 5.2E). The greatest effect was observed in the length of the palatine bone.

5.3.2 Effects on *SMOC2* Transcription

5.3.2.1 Quantitative PCR

LINE-1 retrotransposons are known to be capable of affecting local gene expression through a variety of methods. Mechanisms include the introduction of premature stop codons or polyadenylation signals, providing alternative splice sites and stimulating RNA polymerase disengagement from DNA through transcriptional elongation (Speek 2001; Han *et al.* 2004; Estécio *et al.* 2012). To investigate the potential effects of the *SMOC2* LINE-1 fragment on nearby genes, total RNA extractions from testes samples of nine dogs representing dolichocephalic (x3), mesocephalic (x3) and dolichocephalic (x3) dogs were extracted and analysed. RT-qPCR targeted at both the 5' (exon 2/3) and 3' (exon 10/11) ends of *SMOC2* displayed a comparable dose-dependent down-regulation for each additional copy of the *SMOC2* LINE-1 (Figure 5.3A). Individuals homozygous for the LINE-1 had a ~5-fold reduction in total SMOC2 transcripts compared to their ancestral counterparts. Individuals carrying a single copy of the insertion displayed an intermediary expression profile. RT-qPCR could not confirm the expression levels of the gene-neighbours of *SMOC2*, thrombospondin 2 (*THBS2*) or dishevelled binding agonist of beta catenin 2 (*DACT2*).

5.3.2.2 RNAseq Analysis

To confirm the RT-qPCR findings and to identify any other mRNA expression perturbations, total mRNA of the same nine dogs was submitted for RNAseq analysis. A comparison between homozygous LINE-1 carriers and non-carriers identified a

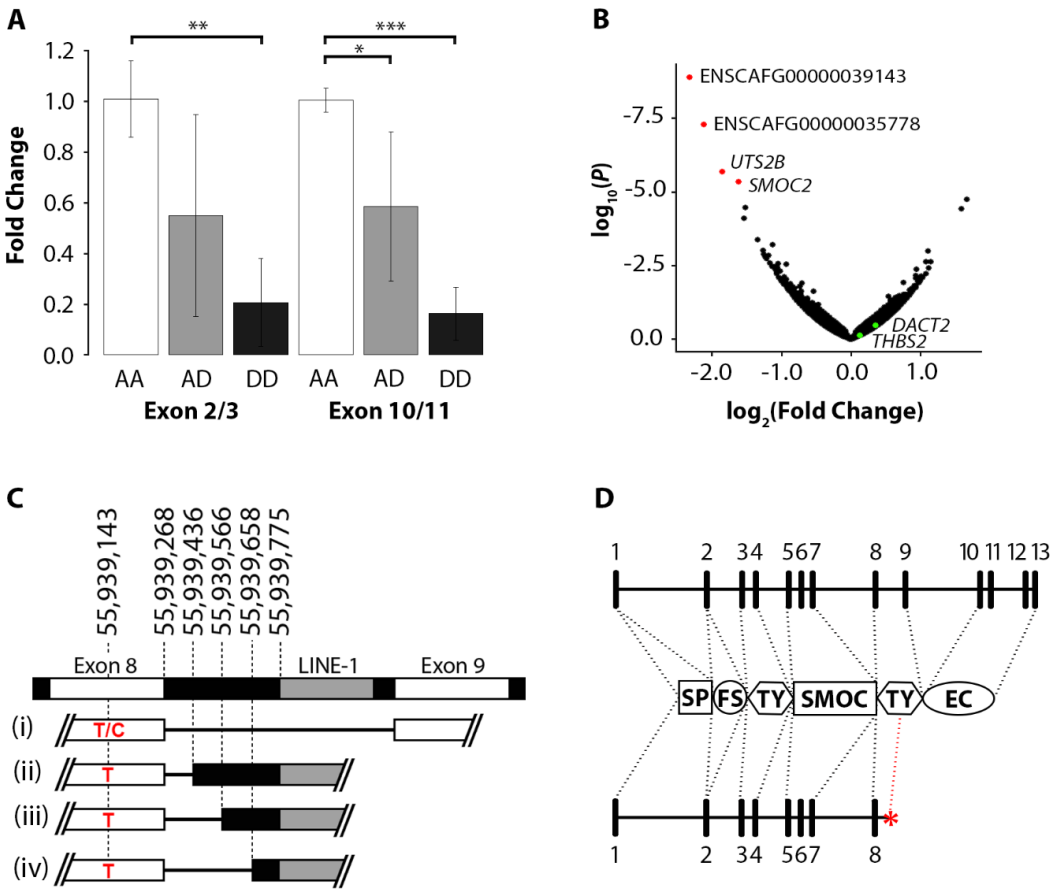


Figure 5.3 – *SMOC2* Transcript Analysis. (A) Efficiency-corrected relative expression of *SMOC2* both up- and downstream of the LINE-1 insertion for individuals homozygous ancestral (AA), heterozygous (AD) and homozygous derived (DD) for the *SMOC2* LINE-1 insertion (<0.05 *; <0.01 **; <0.001 *). (B) RNAseq analysis revealed genes displaying a significant reduction in transcript count in *SMOC2* LINE-1 homozygous carriers compared to non-carriers (red). Genes neighbouring *SMOC2* are coloured green. (C) Schematic diagram of gDNA spanning exons 8 and 9 of SMOC2, including the LINE-1 insertion. mRNA transcripts include the canonical splicing of SMOC2 (i) followed by the three most abundant SMOC2 isofroms when the LINE-1 element is present (ii-iv). All isoforms have premature stop codons prior to exon 9. C/T indicates the SNP in exon 8. (D) Exons 1-13 of *SMOC2* contribute to a follistatin-like module (FS), thyroglobulin-like modules (TY), a unique SMOC module, an extracellular calcium-binding module (EC) and a signal peptide (SP). Modified from: (Marchant *et al.* 2017).**

similar downregulation of *SMOC2* (3.1-fold change) (Figure 5.3B). Three further genes were also identified to have a down-regulation in homozygous carriers which included urotensin 2B (*UTS2B*) and two novel long non-coding RNAs (ENSCAFG00000039143 and ENSCAF00000035778). None of these genes are located on CFA1. No expression changes could be detected in the genes neighbouring *SMOC2*, *THBS2* and *DACT2*.

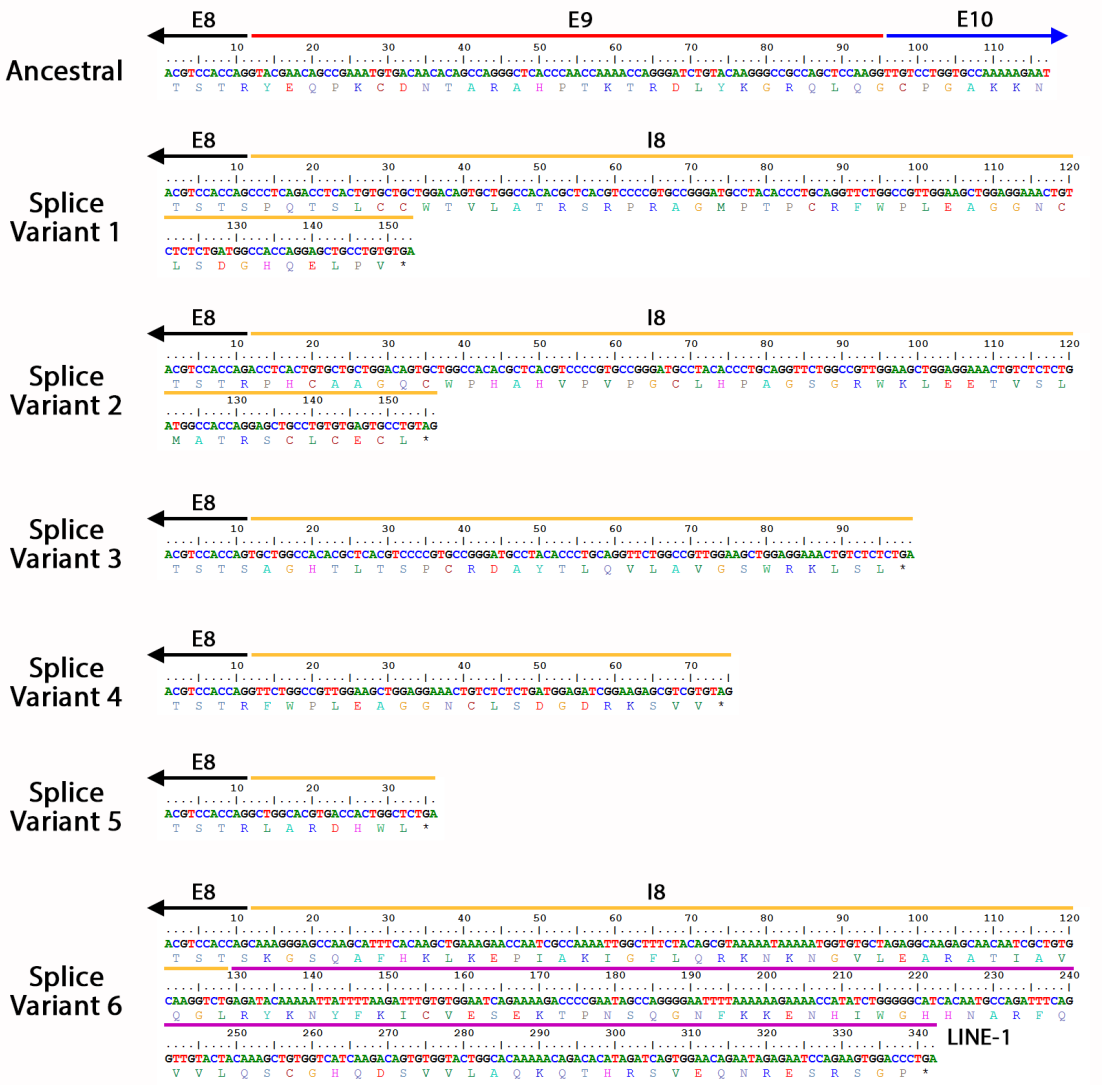


Figure 5.4 – *SMOC2* Isoform Sequences. mRNA and amino acid sequences for the canonical (ancestral) *SMOC2* transcript and the six isoforms identified in dogs that are carriers of the LINE-1 element. Coloured bars indicate the contributions of *SMOC2* exons (E), introns (I) and LINE-1 element into the mature mRNA transcript. All alternative isoforms are predicted to code for premature stop codons (*).Modified from: (Marchant *et al.* 2017).

The assessment of the RNAseq data revealed that all individuals expressed the canonical, 13 exon *SMOC2* transcript, regardless of their LINE-1 genotype status (Figure 5.3C). For homozygous ancestral individuals, this canonical transcript was exclusively expressed however individuals that carried copies of the LINE-1 insertion (either heterozygous or homozygous), expressed multiple different isoforms of *SMOC2*. Using primers targeted against the LINE-1 fragment and exon 8 of *SMOC2*, three abundant splice forms of *SMOC2* across all three individuals homozygous for the LINE-1 insertion were confirmed. Three further rarer transcripts were observed in the RNAseq data of homozygous and heterozygous carriers of the LINE-1 (Figure 5.4). In every instance, the alternative *SMOC2* transcripts have a canonical sequence spanning from exons 1 to 8 however the transcript then splices into a variety of positions within intron 8 prior to reading into the LINE-1 fragment itself. All alternative transcripts characterised are predicted to introduce a premature stop codon following exon 8 (Figure 5.4). It is unclear whether this truncated transcript can produce a mature protein, and if so, what the full effects it may have on its function. However, the protein product is predicted to terminate within the thyroglobulin-like domain and would have no extracellular calcium-binding domain (Figure 5.3D).

The comparison of mRNA and gDNA sequencing revealed that each of the different splice sites within intron 8 of *SMOC2* are preceded by an adenine and guanine residue (AG) – an almost invariant characteristic of mammalian splice acceptors (Breathnach *et al.* 1978; Mount 1982) (Table 5.5).

Transcript	Silent Variant	Splice Acceptor			RNAseq Transcript Count								
		Sequence	Feature	Position	BULD	FBUL	PUG	CKCS	CKCS	PAPI	ITGY	WHIP	YORK
Canonical (Ancestral)	C/T	tcgt**ag** G TAC GAA	Exon 9	55,947,842	224 (66.27)	84 (77.06)	116 (66.67)	258 (95.91)	1012 (95.29)	361 (96.52)	493 (99.60)	644 (100)	911 (99.89)
Splice Variant 1	T	gctc**ag** C CCT CAG	Intron 8	55,939,358	5 (1.48)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Splice Variant 2	T	cctc**ag** A CCT CAC	Intron 8	55,939,365	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (0.53)	0 (0)	0 (0)	0 (0)
Splice Variant 3	T	ggac**ag** T GCT GGC	Intron 8	55,939,387	0 (0)	0 (0)	0 (0)	0 (0)	2 (0.19)	0 (0)	0 (0)	0 (0)	0 (0)
Splice Variant 4	T	ctgc**ag** G TTC TGG	Intron 8	55,939,436	87 (25.74)	17 (15.60)	43 (24.71)	9 (3.35)	40 (3.77)	13 (3.48)	0 (0)	0 (0)	1 (0.11)
Splice Variant 5	T	atac**ag** G CTG GCA	Intron 8	55,939,566	2 (0.89)	1 (0.92)	4 (2.30)	1 (0.37)	3 (0.28)	0 (0)	2 (0.40)	0 (0)	0 (0)
Splice Variant 6	T	tttc**ag** C AAA GGG	Intron 8	55,939,658	20 (5.92)	7 (6.42)	11 (6.32)	1 (0.37)	5 (0.47)	0 (0)	0 (0)	0 (0)	0 (0)

Table 5.5 – Splice Site Sequence for *SMOC2* Isoforms. The seven most frequent transcripts identified which correspond to those in Figure 5.4. The silent variant at chr13:55,939,143 is listed as C or T. Upper-case letters indicate sequence included in mRNA. Lower-case letters indicate intron sequence that is not included in mature mRNA transcripts. Bold letters are the consensus splice donor (gt) and splice acceptors (ag) sites. Three dogs for each *SMOC2* LINE-1 genotype status were used for RNAseq analysis - homozygous derived; Bulldog (BULD), French Bulldog (FBUL) and Pug (PUG); heterozygous Cavalier King Charles Spaniel (CKCS) x2, Papillion (PAPI); homozygous ancestral Italian Greyhound (ITGY), Whippet (WHIP) and Yorkshire Terrier (YORK). The percentage of splice variant read count of the canonical transcript count for each dog are given in brackets. Modified from: (Marchant *et al.* 2017).

5.3.2.3 Allele-specific Expression

Using the RNAseq data, a silent T/C substitution (ENSCAFT00000001348.4:c.699T>C) was observed at position 13:55,939,143 in exon 8 of *SMOC2* (Figure 5.3C). Both the T and C alleles at this locus are observed across individuals that are ancestral for (do not have) the *SMOC2* LINE-1 insertion. Conversely, whenever the *SMOC2* LINE-1 is observed, it is always accompanied by the silent T-allele in exon 8 (Figure 5.3C). This suggests that the T/C silent variant predates the insertion of the *SMOC2* LINE-1 element which occurred into just the T-allele.

This observation permits the inference of allele-specific *SMOC2* expression under certain conditions i.e. whenever the T/C allele is heterozygous. This was the case in three dogs. Firstly, a Yorkshire Terrier (dog 9) was homozygous ancestral (AA) for the *SMOC2* LINE-1 insertion whilst being heterozygous for the T/C allele (Table 5.5). *SMOC2* transcripts from this dog had a percentile ratio of 54:46 for the T and C-allele, respectively suggesting an equal contribution from both alleles to the final *SMOC2* transcript count. This would be expected since there is no LINE-1 insertion present to disrupt transcription.

Dog	Breed	*SMOC2* LINE-1 Genotype	Percentage Transcripts	
			T-allele	C-allele
1	Bulldog	DD	100%	0%
2	French Bulldog	DD	100%	0%
3	Pug	DD	100%	0%
4	Papillion	AD	100%	0%
5	Cavalier King Charles Spaniel	AD	24%	76%
6	Cavalier King Charles Spaniel	AD	25%	75%
7	Italian Greyhound	AA	100%	0%
8	Whippet	AA	100%	0%
9	Yorkshire Terrier	AA	54%	46%

Table 5.6 – Allele-specific Expression of *SMOC2*. RNAseq data was generated for nine dogs for the homozygous derived (DD), heterozygous (AD) and homozygous ancestral (AA) genotypes of *SMOC2*. The sequencing of the C/T silent mutation in exon 8 is given as a percentage of the total read counts for the 55,939,143 base position. The LINE-1 element is found on the T-allele which suggests the decrease in transcript abundance is from the LINE-1 carrying haplotype. Modified from: (Marchant *et al.* 2017).

In contrast, two Cavalier King Charles Spaniels (dog 5 and 6) were heterozygous for both the LINE-1 insertion and the T/C allele (Table 5.5). In both cases, *SMOC2* transcripts had percentile ratios of approximately 25:75 for the T and C alleles, respectively. In both of these incidences, the transcripts harbouring the T-allele are derived from the *SMOC2* LINE-1 containing strand. Therefore, this data suggests that the *SMOC2* LINE-1 insertion is acting in-cis to reduce the total transcript count of the gene. The mechanism for this action is unclear however the LINE-1 element may decrease transcriptional activity, induce nonsense-mediated decay following the production of truncated transcripts, or both.

5.3.3 Species Conservation of *SMOC2*

Morpholino studies using zebrafish suggest that *smoc2* is involved in chondrogenesis of the head (Bloch-Zupan *et al.* 2011; Melvin *et al.* 2013). In order to

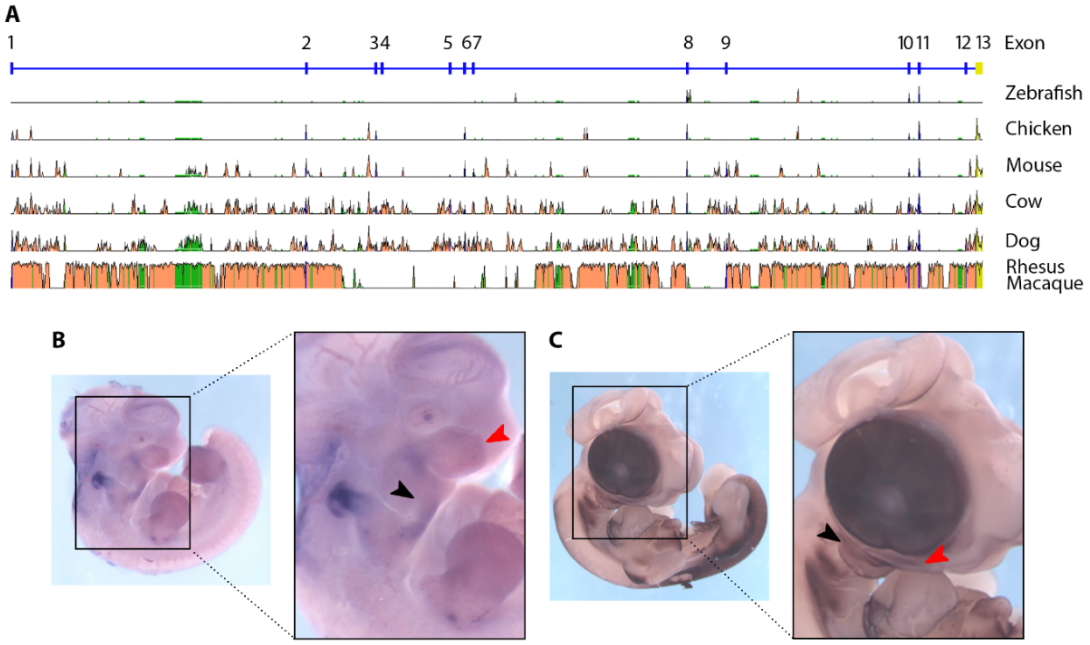


Figure 5.5 – Species Conservation of *SMOC2*. (A) The Evolutionary Conserved Regions of human *SMOC2* spanning chr6:168,841,831 - 169,068,674 (hg19 construct) are poorly conserved in the Zebrafish, Chicken and Mouse compared to the Dog, Cow and Rhesus Monkey. Orange indicates conserved intronic sequence, green indicates conserved repetitive elements. Despite this, *Smoc2* mRNA is expressed in the first pharyngeal arch of the (B) mouse and (C) chicken embryo. *Smoc2* expression is observed in the mandibular process (black arrow) and maxillary process (red arrow) of the first pharyngeal arch. Modified from: (Marchant *et al.* 2017).

investigate the role for *SMOC2* across evolutionary conserved species, the genomic sequence of *SMOC2* across species was compared to humans (Figure 5.5A). This revealed that the Rhesus Macaque had the greatest degree of sequence conservation across the species assessed then followed by the dog and cow. Surprisingly, the zebrafish, chicken and mouse showed a striking reduction in the level of conservation.

Despite this, *Smoc2* mRNA is expressed in the pharyngeal arches of the mouse and chicken embryo (Figure 5.5B&C) (Liu *et al.* 2007). Specifically, the *Smoc2* mRNA

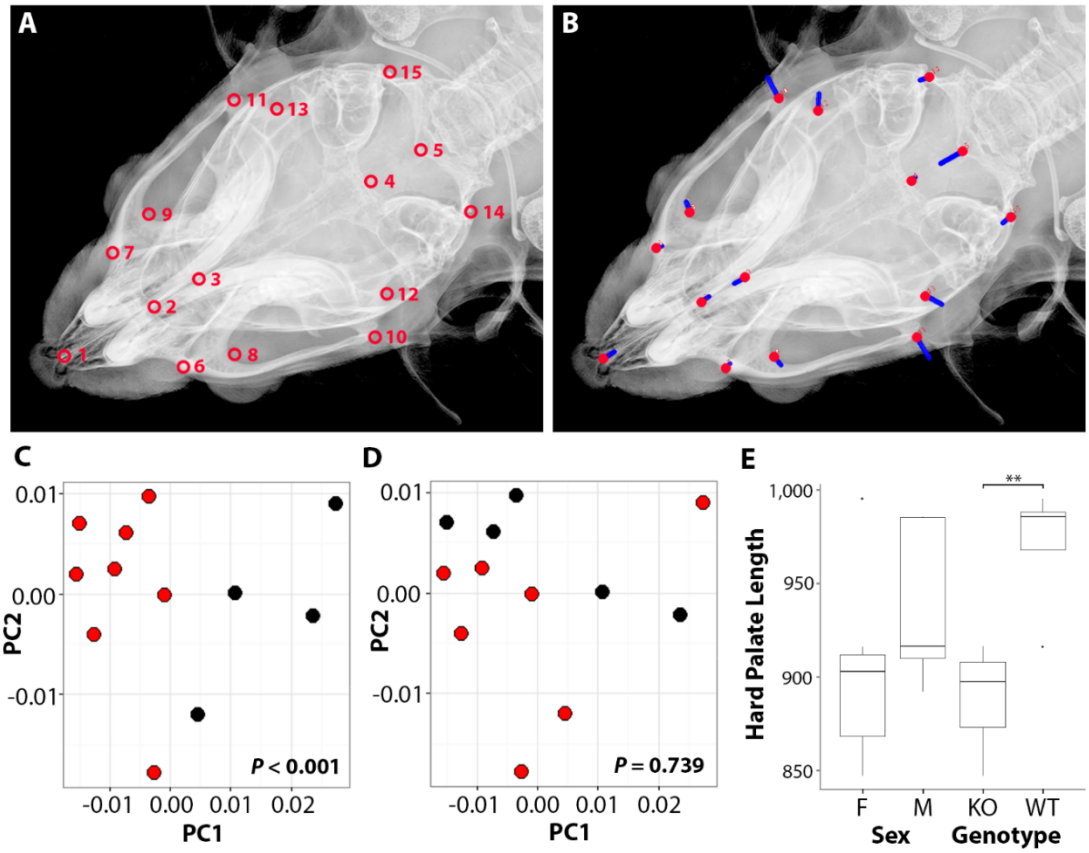


Figure 5.6 - *Smoc2*-null Mice Have Dysmorphic Skulls. (A) Dorsoventral radiographs of mice aged 13 weeks indicating the position of fifteen landmarks across the skull. (B) Lollipop diagram denotes the mean position of landmarks (red circles) and the extent and direction of landmark variation (blue bars) for whole head PC1. The whole skull PC1 values for Smoc2-null (red) and control (black) mice segregate significantly by (C) genotype but not (D) sex. (E) The non-normalised hard palate lengths of Smoc2-null mice are significantly shorter than their WT counterparts whilst no significant difference could be detected between sexes. Modified from: (Marchant *et al.* 2017).

appears to be localised to the mandibular process and the maxillary process of the first pharyngeal arch.

Previous to this study a *Smoc2* null mouse was generated at the Jackson Laboratory, USA and phenotyped as part of the International Mouse Phenotyping Consortium (IMPC). *Smoc2*$^{-/-}$ mice were viable and reached sexual maturity however the lines were no longer kept. As part of the phenotyping screen, radiographs of the skulls were archived. Two-dimensional geometric morphometric analysis was used to assess the variability between the *Smoc2*$^{-/-}$ and wild type (WT) littermates. Fifteen symmetric landmarks captured the variability across the whole skull prior to a Procrustes fit and whole-head principal component (PC) analysis comparable to that carried out in the dog skull reconstruction analysis (Figure 5.6A). Whole-head PC1 accounted for 51.15% of the total variability across the mouse skulls assessed. In the negative direction of whole-head PC1, the length of the skull along the body axis is reduced whilst skull width is widened (Figure 5.6B). This morphological change is comparable to the changes observed in the brachycephalic canine skull. Principal component 1 segregates significantly by *Smoc2* genotype with *Smoc2*$^{-/-}$ mice having shorter and broader (brachycephalic-like) skulls compared to their WT littermates (Figure 5.6C&E). Meanwhile, there was no significant correlation between PC1 and sex (Figure 5.6D&E).

5.4 Discussion

The data presented in this chapter confirms *SMOC2* and its associated LINE-1 insertion as highly compelling candidates for causing canine brachycephaly. Absolute proof of causality could not be shown, which remains a challenge within the field of canine genetics due to the inaccessibility of appropriate tissues, scarcity of cell lines and technical challenges of generating surrogate animal models.

SMOC2 (SPARC-related modular calcium-binding protein 2) is secreted to the extracellular matrix across a wide variety of early embryonic tissues including the heart, lungs, kidneys and developing skeleton (Termine *et al.* 1981; Feng *et al.* 1996; Zhang *et al.* 2008; Maier *et al.* 2008). It belongs to the BM-40 (SPARC) family of

matricellular proteins which are collectively defined by their extracellular calcium-binding domain and a follistatin-like domain. A distinction of *SMOC2* is the addition of two thyroglobulin-like domains and a unique SMOC module (Vannahme *et al.* 2003). It is thought that *SMOC2* is under the control of hedgehog signalling and itself can regulate the expression of bone morphogenetic protein (*BMP*) target genes (Mommaerts *et al.* 2014). Furthermore, due to the interaction of *SMOC2* with cell surface integrins and vascular endothelial growth factor (*VEGF*), platelet-derived growth factor (*PDGF*) and basic fibroblast growth factor (b*FGF*), it is thought to act as a mediator between growth factors and their receptors (Rocnik *et al.* 2006; Liu *et al.* 2007; Maier *et al.* 2008). As a result *SMOC2* has been shown to control cellular adhesion, migration, mitogenesis and differentiation (Rocnik *et al.* 2006; Liu *et al.* 2007; Maier *et al.* 2008; Peeters *et al.* 2018).

Originally, *SMOC2* was first identified in bone where it is known to have a differential expression across the long-bone growth plates (Zhang *et al.* 2008). Mounting evidence suggests that *SMOC2* plays an important role in craniofacial development across species. A zebrafish knockdown study of *smoc2* caused craniofacial hypoplasia which may act, in part through BMP pathways (Melvin *et al.* 2013; Mommaerts *et al.* 2014). *Smoc2* is prominently expressed in the pharyngeal arches of both the developing chicken and mouse embryos. Over the course of embryonic days 10.5 and 12.5 which sees dynamic facial growth in the mouse, *Smoc2* is shown to have a differential expression in the frontonasal and maxillary/mandibular prominences – both of which give rise to craniofacial skeletal tissue (Feng *et al.* 2009). The findings of this thesis further support the importance of *SMOC2* in craniofacial development.

The geometric morphometric analysis of mouse radiographs have shown that *Smoc2*$^{-/-}$ mice have dysmorphic skulls and that they cluster distinctly from their WT littermates. Interestingly, the global effects on skull morphology in the mouse are far less striking than those seen in the domestic dog. An explanation may be that the brachycephalic dog breeds have been heavily selected for their distinct skull

morphology, a phenotype that is known to be complex and likely to have multiple contributing variants that are enriched and reaching fixation in the breeds giving a more apparent extreme phenotype compared to the knockout mouse. Also, the allometric disparity between the mouse and the dog may in some part dilute the phenotypic effects of a *Smoc2* loss of function mouse. Furthermore, challenges of 2D geometric morphometrics makes defining morphological changes in these data more difficult. Namely, variation the positional origin of the image acquisition can alter the apparent positioning of points from one another in 2D compared to 3D image analysis, an affect known as parallax (Mullin and Taylor 2002). Furthermore, 3D skull isosurfaces provide a more reliable, repeatable phenotype capture compared to radiographs which can be harder to interpret. Nevertheless, the *Smoc2*$^{-/-}$ mouse skulls are dysmorphic and a larger follow-up assessment of 3D morphological variation would distil the dysmorphology by shedding light on the regions of the skull most affected by *Smoc2* loss of function.

RNA profiles were analysed in order to investigate the potential effects of the LINE-1 on *SMOC2* biology and function in craniofacial development. In these studies testes were used as a source of total RNA. Embryonic tissues such as the pharyngeal arches or healthy adult craniofacial tissues would have made for a more appropriate source of RNA, however these are unattainable when working with the domestic dog. The decision to use testes was based upon the ease of access and sufficient supply of healthy tissue as well as knowing that *SMOC2* was expressed in the testes of other species (Pazin et al. 2009; Uhlén *et al.* 2015). An assumption was made that the LINE-1 insertion would have a global effect on *SMOC2* expression irrespective of tissue origin. Whilst the LINE-1 was incorporated into *SMOC2* transcripts and a reduction in total read count of the LINE-1 carriers was observed, it remains to be seen whether these effects are consistent across tissues and would need future confirmation.

Interestingly, in addition to *SMOC2*, the RNAseq analysis also highlighted a downregulation of urotensin 2B (*UTS2B*, ENSCAFT00000022282.3, chr34: 23,110,586-23,167,935) and two novel genes for long intergenic noncoding (linc) RNA

(ENSCAFT00000053759.1, chr15:49,296,364-49,338,229; ENSCAFT00000050250.1, chr20:7,300,269-7,309,695). None of these genes are situated on CFA1 so the likelihood of an interaction between the LINE-1 and these genes is low. *UTS2B* is a potent neuropeptide expressed in a variety of tissues including the kidneys, heart and bladder where it is best characterised for its activity as a vasoconstrictor (Matsushita *et al.* 2001; Merlino *et al.* 2013). A link between *UTS2B* and the lead candidate gene, *SMOC2* is unclear. To date, nothing is known about the novel long non-coding RNAs.

The RT-qPCR experiments confirmed a similar downregulation of *SMOC2* in the homozygous carriers of the LINE-1 insertion compared to individuals without a copy (ancestral) with a five-fold change (compared to 3 in the RNAseq data). Through the use of RT-qPCR, a stepwise, downregulation of *SMOC2* was observed such that every copy of the LINE-1 allele would have a dose effect on the total *SMOC2* transcript count. A comparable expression profile of *SMOC2* both up- and downstream of the LINE-1 insertion was also confirmed – previous studies have indicated that there may be a greater number of transcripts containing upstream sequence compared to downstream of the LINE-1 (Han *et al.* 2004). Follow up experiments (discussed in Chapter 7) would be needed to confirm the dose-effect of LINE-1 genotypes on *SMOC2* expression given the standard error in the current RT-qPCR data.

With regards to the precise molecular mechanism of the LINE-1 insertion, the RNAseq data supports evidence for the introduction of multiple splice acceptors in the presence of the LINE-1. All resulting alternative isoforms identified here are predicted to introduce a premature stop codon after exon 8 of the canonical 13 exon *SMOC2* transcript. The resulting hypothesis is that through nonsense-mediated decay, these alternative transcripts are targeted and degraded giving a reduced total *SMOC2* count. No transcripts truncated at exon 8 were identified – all isoforms extended into different positions of intron 8 and the LINE-1 suggesting that a premature polyadenylation and transcription termination is less-likely to be the cause of loss of *SMOC2* expression. Finally, through the identification of a silent C/T variant in exon 8 of *SMOC2*, the LINE-1 in *SMOC2* was predicted to be cis-acting. The

ancestral *SMOC2* allele (without LINE-1) can either have the silent C or T variant whilst every instance of the derived *SMOC2* allele (with LINE-1) has the T-variant in exon 8, suggestive that the exonic variant predates the introduction of the LINE-1 into the region. Through the comparison of the total transcript counts in individuals heterozygous for the silent C/T variant, the allelic contribution in the presence and absence of the LINE-1 insertion could be inferred. Absolute transcript counts were not reported, instead a ratio between alleles which indicates that *SMOC2* ancestral:derived (without LINE-1:with) allelic contributions are approximately 75:25.

Whilst it's believed that the targeted degradation of transcripts is their most plausible fate, it remains unclear as to whether the alternative transcripts may contribute to a mature protein. If a protein product is created it would be predict to terminate partway through the thyroglobulin-like domain and lack its extracellular calcium-binding domain (Vannahme *et al.* 2003). The function of this potentially truncated thyroglobulin domain is unclear however the extracellular calcium binding domain has been shown to be active in mice and to instigate a conformational change of the protein when bound to a calcium ion (Vannahme *et al.* 2003). The absence of this domain would undoubtedly have an effect on the hypothetical protein product function, although the precise nature of which is unclear. This function may act in a dominant negative fashion but further experimentation would be needed to assess this.

In leveraging the diverse craniofacial form of dogs, *SMOC2* and its associated LINE-1 element are believed to dictate the brachycephalic phenotype. Hopefully, this information can be selectively used to further explore craniofacial anomalies and associated diseases in both dogs and humans.

Chapter 6: Characterisation of Upper Airway Syndrome in Norwich Terriers

6.1 Introduction

In the previous chapter, the brachycephaly-associated LINE-1 in *SMOC2* was characterised in an attempt to understand more about the contributors to craniofacial variation in the dog. The confounding effects of the brachycephalic skull shape have long been considered the major predisposing factor to BOAS however heterogeneity in disease presentation both across and within breeds suggest that other factors may also be involved. To assist in deciphering disease aetiology, the Norwich Terrier became of interest.

The Norwich Terrier is not considered to have a classically brachycephalic skull shape and yet it presents with Upper Airway Syndrome (UAS), a disease which has many of the same hallmarks as BOAS. These hallmarks include an elongated and thickened soft palate, everted laryngeal saccules and oedema of the upper airways (Koch *et al.* 2003; Torrez and Hunt 2006; Dietschi *et al.* 2010; White 2011; Johnson *et al.* 2013). In this chapter, the cause of UAS in the Norwich Terrier has been determined which has important ramifications for the management of the condition in Norwich Terrier as well as providing new perceptions for BOAS susceptibility in brachycephalic breeds.

All work contained within Section 6.2 is a word-for-word copy of a manuscript recently submitted to the journal, PLoS Genetics (submitted: 23 July 2018). All analyses described were performed by Thomas Marchant with the exception of performing and grading endoscopic examinations (Dr Elisabeth Dietschi, Dr Ulrich Rytz and Dr Peter Schawalder), extraction of DNA and assistance with Sanger sequencing (Sheida Hadji Rasouliha). The manuscript was prepared by Thomas Marchant with contributions from Dr Jeffrey Schoenebeck, Dr Dylan Clements and Prof Tosso Leeb. All figures were generated by Thomas Marchant. Affiliations of all contributors are given in the manuscript. All supplementary materials for the

submitted manuscript have been added in Appendix 1. All references cited in this manuscript are listed in Section 6.2.9 using the journal's Vancouver style. Additional methods, results and discussions that were not included in the manuscript have been written specifically for this thesis and have been added in Sections 6.3 and 4.4.

6.2 Submitted Manuscript

6.2.1 Title Page

An *ADAMTS3* Missense Variant is Associated with Norwich Terrier Upper Airway Syndrome

Thomas W. Marchant[1], Elisabeth Dietschi[2], Ulrich Rytz[3], Peter Schawalder[3], Sheida Hadji Rasouliha[2], Ronan S. Harrington[1], Michaela Drögemüller[2], Jeffrey Kidd[4], Elaine A. Ostrander[5], Amanda Warr[1], Mick Watson[1], David Argyle[1], Gert Ter Haar[6†], Dylan N. Clements[1], Tosso Leeb[2] & Jeffrey J. Schoenebeck[1*]

1. The Roslin Institute and Royal (Dick) School for Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, UK.
2. Institute of Genetics, Vetsuisse Faculty, University of Bern, 3001 Bern, Switzerland.
3. Department of Clinical Veterinary Medicine, Division of Small Animal Surgery, Vetsuisse Faculty, University of Bern, 3001 Bern, Switzerland.
4. Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA.
5. Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.
6. Department of Clinical Sciences and Services, Royal Veterinary College, Hertfordshire AL9 7TA, UK

[†]GTH current contact details: Specialistische Dierenkliniek Utrecht, Middenwetering 19, Utrecht, 3543 AR, NE

*corresponding author: jeff.schoenebeck@roslin.ed.ac.uk

6.2.2 Abstract

In flat-faced dog breeds, air resistance caused by skull conformation is believed to be a major determinant of brachycephalic obstructive airway syndrome. The clinical presentation of BOAS is heterogeneous, suggesting determinants independent of skull conformation contribute to airway disease. Norwich Terriers, a mesaticephalic breed, are predisposed to Upper Airway Syndrome (UAS), a disease whose pathological features overlap with BOAS. Our health screening clinic examined and scored the airways of 401 Norwich terriers by laryngoscopy. Genome-wide association analyses of UAS-related pathologies revealed a genetic association on canine chromosome 13 (rs9043975, p=7.79x10^{-16}). Whole genome resequencing was used to identify causal variant(s) within a 414 kb critical interval. This approach highlighted an error in the CanFam3.1 dog assembly, which when resolved, led to the discovery of a c.2786G>A missense variant in exon 20 of the positional candidate gene, a disintegrin and metalloproteinase with thrombospondin motifs 3 (*ADAMTS3*). In addition to segregating with UAS amongst Norwich Terriers, the *ADAMTS3* c.2786G>A risk allele frequency was enriched among the BOAS-susceptible French and (English) Bulldogs. Previous studies indicate that *ADAMTS3* loss of function results in lymphoedema. Our results suggest a new paradigm in the understanding of canine upper airway disease aetiology: airway oedema caused by disruption of *ADAMTS3* predisposes dogs to respiratory obstruction. These findings will enhance breeding practices and could refine the prognostics of surgical interventions that are often used to treat airway obstruction.

6.2.3 Author Summary

Respiratory diseases are prevalent across the dog, particularly in brachycephalic breeds such as the Bulldog and French bulldog. The flat facial conformation of these breeds have long been assumed to be the major predisposing factor however the underlying genetics of their respiratory condition has never been elucidated. Here, we became interested in the Norwich terrier, a breed presenting with many of the same respiratory disease symptoms as the Bulldog. A distinction,

however, is that the Norwich terrier is not considered to be a brachycephalic breed and so presented an opportunity to disseminate respiratory disease from head conformation. We performed a genome-wide association analysis for respiratory disease severity in the Norwich terrier and resolved an association on chromosome 13 to a missense mutation in *ADAMTS3*. Variation in this gene have been previously shown to cause an oedematous phenotype – a disease characteristic in the airways of Norwich terriers and brachycephalic dogs alike. We screened over 100 breeds for the *ADAMTS3* variant and found it to be enriched in the Norwich terrier, Bulldog and French bulldog. This discovery changes how we view respiratory disease predisposition in the dog, offers potential genetic screens and highlights a new biological function for *ADAMTS3*.

6.2.4 Introduction

Among dogs, brachycephaly describes the head conformation of many popular breeds including the Bulldog, French Bulldog and Pug. This trait is grossly characterised by the concurrent rostrocaudal shortening and mediolateral widening of the skull and is accompanied by skin folds of the face. The structural discordance between the reduced facial skeleton and its overlying soft tissues such as the wrinkled skin folds underpins these breeds' iconic looks, but these artificially selected aesthetics are under increasing scrutiny for their association with health problems including breathing difficulties.

It is thought that soft tissues of the upper respiratory tract such as the nostrils, nasal turbinates and soft palate do not scale proportionately with reductions in the midface skeleton [1]. Misconfiguration of respiratory soft tissue restricts airflow and increases negative pressure within the airway [1,2]. This predisposes brachycephalic dogs to Brachycephalic Obstructive Airway Syndrome (BOAS). Dogs diagnosed with BOAS can have stenotic nares, elongated soft palates and oversized, caudally protruding nasal turbinates [2-7]. Airway resistance caused by these soft tissue anomalies is believed to induce pathological remodelling of additional tissues including tonsil and laryngeal saccule eversion, oedema of the nasopharynx,

laryngeal collapse, tracheal hypoplasia and exacerbation of the thickening and elongation of the soft palate [2,8,9]. Collectively, these perturbations severely impact the wellbeing of affected individuals by increasing their respiratory effort, resulting in laboured breathing, intolerance to heat/exercise, cyanosis and collapse [6,7].

The clinical assessment of the respiratory obstruction is often based on the grading of clinical symptoms, diagnostic imaging, and more recently, whole-body barometric plethysmography [6,10,11]. Treatment options for BOAS include anti-inflammatory medication which can reduce swelling/oedema acutely, however corrective surgery is often required to alleviate the condition [12]. Rhinoplasty of the nares, excision of the caudal aspect of the soft palate and aberrant turbinates, removal of the laryngeal saccules and tonsillectomy are the most common surgical procedures, which generally have mixed prognoses [2,3,6,12-14]. The number of patients requiring surgical treatment is expected to rise notably with the rapid increase in popularity of brachycephalic breeds. The costs and morbidity of surgical treatment are a welfare concern for both owners and their dogs, with complications reported in up to 25% and mortality in as many as 5% of cases treated surgically [14].

We and others have studied the underlying genetics of canine skull shape variation [15-19]. Variants in the *BMP3* and *SMOC2* genes are associated with canine brachycephaly, however the contribution of these variants to BOAS pathogenesis is unclear. Moreover, variants in both of these genes appear largely fixed among brachycephalic breeds that are at greatest risk of developing BOAS and yet the incidence and severity of BOAS differs between them [11,20]. BOAS heterogeneity may also be influenced by environmental and epigenetic factors, as well as other genetic modifiers segregating among dog populations. Under this premise, we became interested in the presentation of a respiratory condition remarkably similar to BOAS which has been identified in Norwich Terriers.

As their name suggests, Norwich Terriers originate from south eastern UK where they were used for rodent control. Today, Norwich Terriers are recognised by all major kennel clubs and are known for their short, stocky build and prick ears.

Dietschi *et al.* first described the presentation of Upper Airway Syndrome (UAS) in the Norwich Terrier. Although they are not considered a brachycephalic breed, affected Norwich Terriers present many of the hallmarks of BOAS including elongated and thickened soft palates, oedema of the nasopharynx and everted laryngeal saccules [21,22]. The closely related Norfolk Terrier, a breed that officially split from the Norwich Terrier in 1964, is seemingly unaffected by UAS, suggesting genetic predisposition in the Norwich Terriers

Moreover, anecdotes from breeders regarding more recent dog generations, suggested that some Norwich Terriers appeared shorter-faced than those from earlier generations (personal communication to JS). Indeed, Koch *et al.* postulated that selective breeding is driving Norwich Terriers to become brachycephalic [9]. Spurred on by these observations, and the possibility of uncovering genetic modifiers that increase respiratory obstruction risk, we sought to understand the genetic basis of Norwich Terrier UAS.

6.2.5 Methods

6.2.5.1 Participants and Upper Airway Assessment

All animal experiments were performed according to the local regulations. The dogs in this study were examined with the consent of their owners. The study was approved by the "Cantonal Committee for Animal Experiments" (Canton of Bern; permits 22/07, 23/10, and 75/16) and the R(D)SVS Veterinary Ethical Review Committee (20 16, University of Edinburgh). A full description of the upper assessment was described previously [22]. In short, the upper respiratory tracts of 401 Norwich Terriers and 12 Norfolk Terriers were assessed *in situ* during endoscopic examination and scored retrospectively from video footage. All evaluations were conducted by a single veterinary surgeon. Subsequently, each of the ten phenotypic components of the airway (soft palate length, soft plate thickness, laryngeal saccule, cartilage shape, cartilage stability, cartilage position, oedema of the oropharynx, oedema of the pharynx, oedema of the cricoid mucosa and shape of the trachea) were scored on a range from 1 to 4 representing 'normal', 'mild', 'moderate' and

'severe' respectively by authors PS, ED and UR. A custom R script was used to generate Pearson's correlation and dendrogram. Individual phenotype scores were weighted and summed to give the total airway score.

Two-hundred and thirty-three Norwich Terriers (109 male, 124 female) representing the extremes of upper airway phenotypes formed the study cohort. Participants varied in age from 7 to 188 months (median = 18 months).

6.2.5.2 Skull Shape Assessment

The geometric morphometric analysis of 3D skull reconstructions generated from computer tomography scans have been described previously [18]. Linear measurements of the hard palate were made, and the influence of allometry regressed using the neurocranium centroid size. Principal component analysis of the viscerocranium of 565 dogs representing 96 UK Kennel Club registered breeds permitted the comparison of face shapes.

6.2.5.3 Genotyping and Genomic Analysis

Whole blood samples were taken and stored in EDTA at 4 °C prior to gDNA extraction following the whole blood protocol of the Nucelon BACC Genomic DNA Extraction Kit (RPN-8502, GE Healthcare Life Sciences). Genotypes were generated using the Illumina 170,000 SNV CanineHD bead chip by Edinburgh Genomics, UK and mapped to the CanFam3.1 coordinates.

SNVs with minor allele frequencies < 0.05 and individuals with missingness > 0.1 were removed using PLINK (v1. 90) [52]. Genotypes were imputed using a two-step process that included pre-phasing by SHAPEIT [53] and imputation by IMPUTE2 [54]. A total of 105,130 SNVs were used by GEMMA (v0.94.1) in a univariate linear mixed model to perform genome-wide association analyses [55]. A kinship matrix was implemented during the analysis with age and sex used as covariates. A Bonferroni correction threshold was used to determine statistically significant SNVs ($-\log_{10}$ [0.05/105,130] = 6.32). The linkage disequilibrium of significant SNVs with all

other markers in a 50 variant window was calculated using the independent pairwise test in PLINK (v1.90).

Phased haplotypes encompassing the index SNV (TIGRP2P185081_rs9043975) at chr13: 61,255,943 were ordered by the UAS severity. The order was dictated by the four phenotypes returning significantly associated index SNVs (laryngeal saccule > cartilage position > oedema of the cricoid mucosa > oedema of the oropharynx) such that dogs with the most severe grade of all four phenotypes were positioned at the top. A consensus risk haplotype was the most frequent haplotype within the most severe scoring dogs, appearing in 98 of 162 chromosomes from severely affected dogs. Risk alleles were coloured based on whether they matched this consensus haplotype. The critical interval boundaries were defined by three or more meiotic recombination events across the severely affected Norwich Terriers.

6.2.5.4 Sequencing and Variant Analysis

Four Norwich Terriers representing upper airway phenotypic extremes were whole genome sequenced to an average coverage of 15.9x. Two Norwich Terriers were homozygous for the disease-associated haplotype with one severely affected and the second apparently unaffected. The remaining two dogs did not have the disease-associated haplotype and were unaffected. DNA libraries were prepared using the TruSeq DNA PCR-free Library Preparation Kit. The Illumina HiSeq 2500 system sequenced 125 bp paired-end libraries with and average insert size of 419 bp. Reads from each resequenced Norwich Terrier were aligned to CanFam3.1 assembly using BWA-MEM [56] and variants within the critical interval chr13:61,166,179 - 61,579,985 were called for the CanFam3.1 and Zoey2.3 assembly using Platypus (v0.8.1) [57]. Two Norwich Terriers homozygous for the disease-associated haplotype with differing phenotypes were selected with the potential of discovering the ancestral haplotype prior to the introduction of the causal variant(s). To this end, for positions that had calls for all four Norwich Terriers, filtering criteria required variant(s) to be homozygous and exclusive to the affected dog however this returned no variants. We hypothesised that the unaffected dog homozygous for the disease-

associated haplotype was still subclinical due to the age of scoping at 1.2 years. Subsequently, filtering criteria required variants to be homozygous derived in the disease-haplotype carrying dogs and homozygous ancestral in those not carrying it.

Norwich Terrier DNA and CanFam3.1-aligned RNAseq reads were viewed in IGV [58] which revealed an error in the reference sequence [18]. To resolve this error, we compared the local assembly of a Great Dane produced from PacBio long reads (unpublished data, author JK). In addition, we generated a *de novo* assembly from three bacterial artificial chromosomes (BACs) originally used for the CanFam3.1 assembly. BACs spanning the critical interval were sourced from the BACPAC Resource Center, Children's Hospital Oakland Research Institute, California, USA (CH82-24F19, CH82-379O18 and CH82-101M10). Following BAC isolation (PhasePrep™ DNA Kit, Sigma-Aldrich, NA0100), a DNA library was prepared from an equimolar mix of the three BACs for a single 1D barcode-free gDNA sequencing run using the Oxford Nanopore Technologies MinION platform (SQK-LSK109, R9.4). A pipeline including Albacore (v2.0.1), Canu (v1.5) and Nanopolish (v0.8.4) was used to base call, construct contigs and improve consensus sequence respectively. The consensus sequences of both long-read platforms were in agreement and resolved the error underlying exon 20 of *ADAMTS3*, though neither platform's base calling resolved a ~40 bp intronic stretch of G's downstream of exon 20.

Norwich Terrier short-read and RNAseq data from a previous study (European Nucleotide Archive study accession PRJEB17926) were realigned to the new consensus using BWA-MEM and STAR (V2.5.1) respectively with default parameters [18,59]. The RNAseq data was used solely to confirm exonic structure whilst the DNAseq data was used to repeat variant calling as previously described.

Protein sequences of *ADAMTS3* homologs (HGNC:219) across species were downloaded from Ensembl and aligned using a ClustalW multiple alignment [60]. A low quality canine *ADAMTS3* transcript (XP_539311) was identified however its predicted amino acid sequence surrounding the exon 20 missense variant was substantially different from any other species – likely due to 133 bp of exon 20

missing from the CanFam3.1 assembly. To this end, the predicted amino acid sequence for the Norwich Terrier was created using comparative RNAseq alignment from 9 dogs, representing 8 breeds which were in full agreement for exon structure [18]. Residue positions are relative to the start codon. The thrombospondin-like domain (PS50092) was predicted using PROSITE database of protein domains [61].

To genotype the *ADAMTS3*:c.2786G>A variant, forward (ACACACGAACCCAGGCACAC) and reverse (GGCCTGGGAGCACTGCAC) primers were designed to amplify the region. PCR products were Sanger sequenced by Edinburgh Genomics, UK. All breeds used for genotyping were owner-reported.

6.2.6 Results

6.2.6.1 Distinct Skull Morphology

There is a remarkable overlap in the clinical presentations of respiratory tract disorders across several breeds including the Bulldog, French Bulldog, Pug, Shih Tzu and Norwich Terrier. Despite this, the Norwich Terrier is not considered to be a brachycephalic breed and is morphologically distinct from the other aforementioned

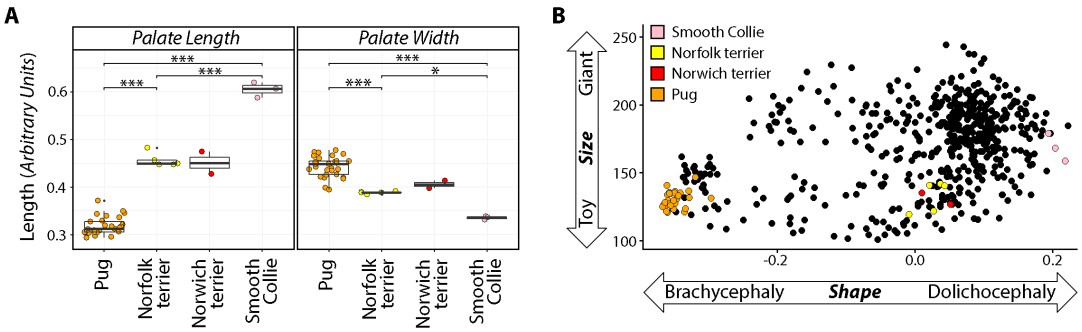


Figure 6.1 - Morphology and Respiratory Distress. (A) Standardised measurements of the canine hard palate infer an intermediate rostrum shape for the Norwich and Norfolk Terriers between the extreme brachycephalic (Pug) and dolichocephalic (Smooth Collie). Mann-Whitney-Wilcoxon and Kolmogorov-Smirnov tests <0.05 *; <0.01 **; <0.001 *. (B) Geometric morphometric analysis of 96 domestic dog breeds reveals that the Norwich and Norfolk Terriers occupy a distinct morphospace from the brachycephalic breeds, such as the Pug that are susceptible to BOAS. Size is scored by neurocranium centroid size (y-axis) and face shape is scored by viscerocranium PC1 (x-axis).**

breeds. Linear measurements of the Norwich Terrier hard palate revealed intermediate palate dimensions between the extremes of facial morphology represented by the brachycephalic Pug and dolichocephalic Collie (Figure 6.1A). Furthermore, geometric morphometric analysis of the canine rostrum revealed that the Norwich Terrier occupies a morphospace distinct from classic brachycephalic breeds such as the Pug (Figure 6.1B). For both linear measurements and geometric morphometrics, our data do not indicate a gross morphological difference between the Norwich and Norfolk Terrier.

6.2.6.2 Norwich Terrier Upper Airway Assessment

In 2000, an upper airway screening programme for Norwich Terriers was established at the Vetsuisse Faculty of the University of Bern in Switzerland. The programme uses laryngoscopic videos to score ten components of the airway which are combined to give an overall grade of the condition [22]. To date, 401 Norwich Terriers in addition to 12 Norfolk Terriers were screened. Two-thirds (65.8%) of Norwich Terriers had clinical presentations of UAS ranging from 'mild' to 'severe' whilst all Norfolk Terriers were unaffected (Figure 6.2A). We selected 233 Norwich Terriers (109 male, 124 female) representing phenotypic extremes of the phenotypic distribution for our study (Figure 6.2B). Everted laryngeal saccules (81, 35%) and elongated soft palates (33, 14%) were the most common severely-graded phenotype and were only graded as normal in 11 (5%) and 43 (19%) dogs respectively (Table 6.1). Meanwhile, cartilage stability (125, 54%) and trachea shape (112, 48%) were the most common normal-graded phenotypes. All unique phenotype pairings display a positive Pearson's correlation coefficient (range: 0.018 to 0.720, median: 0.344) with the exception of the cartilage shape and oedema of the pharynx phenotypes (r = -0.116) (Supplementary Figure 1). Cartilage position and cartilage stability had the highest correlation of any phenotype pair (r = 0.720).

Between 2003 and 2007, The Swiss Terrier Club discouraged breeding dogs exceeding a 'moderate' upper airway phenotype. From 2007 onwards, upper airway

screening was mandatory for all breeding pairs in Switzerland. As a testament to the coordinated efforts between veterinarians and breeders, the Swiss screening programme observed a reduction in the number of severely affected Norwich Terriers from 44.0% of those born between 1988-1997 to 8.6% for those born in 2006-2014 (Figure 6.2C). The success of the screening programme underscores the heritability of UAS and suggests that the disease indeed segregates within this population. With cases of UAS reported across continents, the need to develop portable, cost-effective screening strategies became imperative. In order to further reduce the disease prevalence across the Norwich Terrier population and to provide insights into the pathophysiology of respiratory diseases that affect the upper airways of dogs, we sought to establish the genetic underpinnings of the condition.

Upper Airway Phenotype	Normal	Mild	Moderate	Severe
Laryngeal Saccule Score	11 (4.7%)	75 (32.2%)	66 (28.3%)	81 (34.8%)
Soft Palate Length	43 (18.5%)	67 (28.8%)	90 (38.6%)	33 (14.2%)
Oedema of the Cricoid Mucosa	5 (2.1%)	105 (45.1%)	91 (39.1%)	32 (13.7%)
Oedema of the Pharynx	8 (3.4%)	90 (38.6%)	104 (44.6%)	31 (13.3%)
Oedema of the Oropharynx	41 (17.6%)	128 (54.9%)	56 (24%)	8 (3.4%)
Trachea Shape	119 (51.1%)	87 (37.3%)	21 (9%)	6 (2.6%)
Cartilage Position	97 (41.6%)	92 (39.5%)	39 (16.7%)	5 (2.1%)
Cartilage Stability	125 (53.6%)	76 (32.6%)	30 (12.9%)	2 (0.9%)
Soft Palate Thickness	79 (33.9%)	98 (42.1%)	54 (23.2%)	2 (0.9%)
Cartilage Shape	112 (48.1%)	0 (0%)	121 (51.9%)	0 (0%)

Table 6.1 Norwich Terrier Phenotype Scores. Each upper airway phenotype is graded as 'normal', 'mild', 'moderate' or 'severe' following the assessment of endoscopy videos. Two-hundred and thirty-three Norwich Terriers were assessed with counts for each phenotype and severity grading (percentage across phenotype).

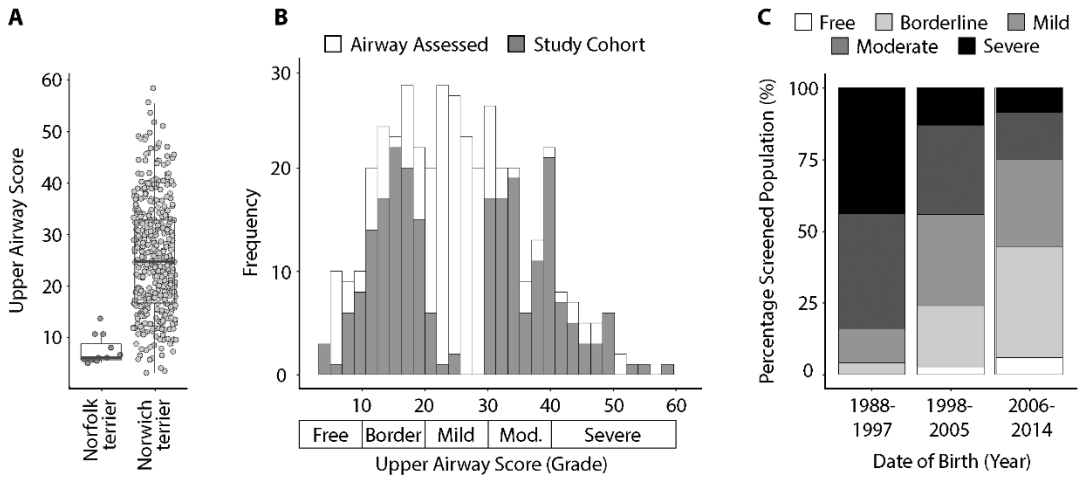


Figure 6.2 - Upper Airway Syndrome across Norwich Terriers. (A) All Norfolk Terrier upper airway scores were considered clear (n=12), whilst 277 Norwich Terriers (n=416) were diagnosed with clinical Upper Airway Syndrome. (B) Two-hundred and thirty-three Norwich Terriers representing the extremes of those screened made up the study cohort. (C) Breeding recommendations based on upper airway scores have reduced the percentage of severely affected Norwich Terriers over time.

6.2.6.3 Genome-wide Association Analysis

Genome-wide association analyses (GWAS) were performed for each of the ten upper airway phenotypes. Four phenotypes including eversion of the laryngeal saccule, oedema of the cricoid mucosa, oedema of the oropharynx and cartilage position returned markers with genome-wide significance (Figure 6.3A, Supplementary Figure 2). The threshold for genome-wide significance was established by Bonferroni correction ($-\log_{10}$ [0.05/105,130] = 6.32). Regardless of phenotype, all association tests highlighted the same ~2.9 Mb quantitative trait locus (QTL) spanning 58,941,974 – 61,830,084 bp on canine chromosome (CFA) 13 with an index marker (TIGRP2P185081_rs9043975) at chr13:61,255,943 (Figure 6.3B, Supplementary Table 6.1). Markers within this broad QTL display high levels of linkage disequilibrium ($r^2 > 0$ .2). Due to the modest correlation between individual traits (Supplementary Figure 1), many markers (35/57) are significantly associated with at least two phenotypes (Supplementary Table 1). Principal components analysis of genotypes did not reveal phenotype-related substructure within the study cohort,

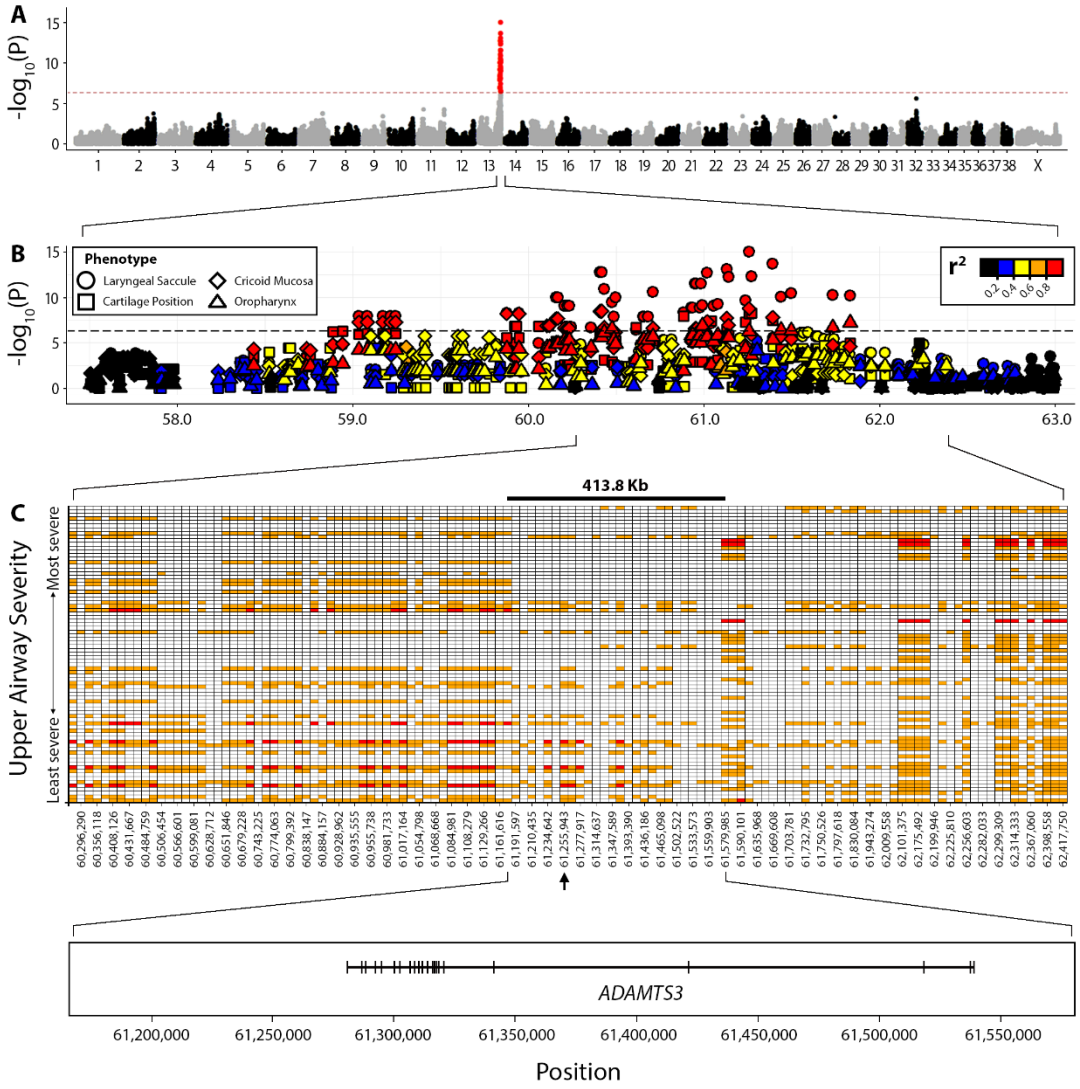


Figure 6.3 - Refinement of CFA13 Critical Interval. (A) The CFA13 QTL was identified across four upper airway phenotypes including laryngeal saccule score. (B) Marker associations for all phenotypes which returned at least one association surpassing the Bonferroni correction threshold (4.75×10^{-7}) are overlaid. Point shapes represent phenotypes whilst colour indicates the degree of linkage disequilibrium (r^2) with significant markers. (C) Genotypes were phased for individual Norwich Terriers (horizontal bars) and ordered by the upper airway phenotypes. Only severely affected dogs are shown. Alleles are coloured white, orange and red for homozygous consensus for the risk allele, heterozygous and homozygous alternate alleles, respectively. A 413.8 kb critical interval was defined by at least 3 meiotic recombination events. The index marker (chr13:61,255,943) is identified by an arrowhead. The critical interval encompasses the *ADAMTS3* gene.

adding confidence that the signal on CFA13 was truly associated with the disease (Supplementary Figure 3).

6.2.6.4 Critical Interval Refinement

Genotypes extending ~1 Mb in both directions from the genome-wide significant markers on CFA13 were phased. Individual dogs were ranked by their disease severity and critical interval boundaries were defined by three meiotic recombinations. This revealed a 413 kb haplotype spanning chr13:61,166,179 – 61,579,985 to be shared across most severely affected Norwich Terriers (Figure 6.3C). The disease-associated haplotype was homozygous in 75.3% (61/81) of severely affected dogs whilst it was homozygous in just 18.4% (28/152) of moderately-to-unaffected dogs (Supplementary Figure 4). This critical interval spans the entirety of the disintegrin and metalloproteinase with thrombospondin motifs 3 (*ADAMTS3*) gene, in addition to ~114 kb and ~41 kb of sequence up and downstream of the gene, respectively. No other protein coding genes were annotated within the critical interval.

6.2.6.5 Identifying Candidate Causal Variants

To search for putative causal variants, we whole genome sequenced four Norwich Terriers representing the extremes of UAS phenotypes. This included two dogs that were homozygous for the CFA13 risk haplotype – one severely affected by UAS and the second seemingly unaffected. The remaining two dogs did not carry the CFA13 risk haplotype and were clinically unaffected. A total of 2,276 variants were called within the 413,806 bp critical interval and subsequently filtered (see Methods), however no variants (SNVs or indels) were compelling candidates for causality based on location and/or interspecies conservation (Table 6.2, Supplementary Table 2). Following visual inspection of the whole genome sequences alongside aligned RNAseq data from a previous study [18], we observed a gap in short-read coverage across all DNAseq and RNAseq datasets at exon 20 of *ADAMTS3* suggesting an error in the CanFam3.1 assembly (Supplementary Figure 5A). We elected to generate a new

	CanFam3.1	New Consensus
Interval (bp)	61,166,179 - 61,579,985	1,164,985 - 1,578,005
Size (bp)	413,806	413,020
Called	2,276	1,834
Passed Filters	231	80
Intronic	234	78
Exonic	1	2
Protein Changing	0	1

Table 6.2 Variant Calling. Summary of the variants called across the four resequenced Norwich Terriers using Platypus within the critical intervals of the (A) CanFam3.1 and (B) newly created consensus sequence. Variants are filtered by the presence/absence of the risk haplotype.

local assembly for the CFA13 critical interval using long-read sequencing (see Methods). DNA- and RNA-seq short reads were aligned to the new consensus sequence and revealed that exon 20 of *ADAMTS3* extended an additional 133 bases beyond what was present in CanFam3.1 (Supplementary Figure 5B). Subsequent variant calling of the new 413,020 bp critical interval identified 1,834 variants. Variants were filtered based on allelic segregation between the disease-associated and alternate haplotypes, leaving a total of 80 single nucleotide variants (SNVs) and small indels. All remaining variants are in complete linkage disequilibrium ($r^2 = 1$). Two of the remaining variants are exonic – a synonymous variant in exon 21 and a missense variant in the newly defined exon 20 (c.2786G>A) (Supplementary Table 2). The missense variant is predicted to change an amino acid of *ADAMTS3* from an arginine to histidine, p.(Arg929His). This arginine is positioned within a thrombospondin-like domain and is invariable across mammalian species and close gene paralogs, suggesting evolutionary constraint (Figure 6.4). Accordingly, a substitution at this position is predicted to be “probably damaging” and “not tolerated” by PolyPhen-2 and SIFT, respectively [23,24].

6.2.6.6 Genotype-Phenotype Association

We genotyped all Norwich and Norfolk Terriers screened in the study for the c.2789G>A variant and did not observe the allele among the Norfolk Terrier

population (n=12), as expected since UAS was not diagnosed in dogs of this breed. However, the risk allele was homozygous in 132 (32.9%) and heterozygous in 195 (48.6%) individuals from the Norwich Terrier cohort (n=401). Dogs homozygous for the c.2786G>A allele had a significantly greater total upper airway score than those heterozygous ($p = 9.10 \times 10^{-17}$) or homozygous ($p = 3.08 \times 10^{-20}$) for the ancestral allele (Figure 6.5A). Seventeen Norwich Terriers were seemingly unaffected by UAS, two of which were homozygous for the c.2786G>A risk allele and nine were heterozygous (Figure 6.5A). Of note, many of the unaffected Norwich Terriers were young (range 11 to 39, median: 14 months old) at the time they were screened. In contrast, the *ADAMTS3* genotype does not segregate with weight (Figure 6.5B). Interestingly, by applying the Swiss Norwich Terrier club breeding guidelines to all 401 screened Norwich Terriers, 74.1% of those prevented from breeding are homozygous for the variant, whereas only 22.0% of Norwich Terriers permitted to breed have this genotype (Figure 6.5C).

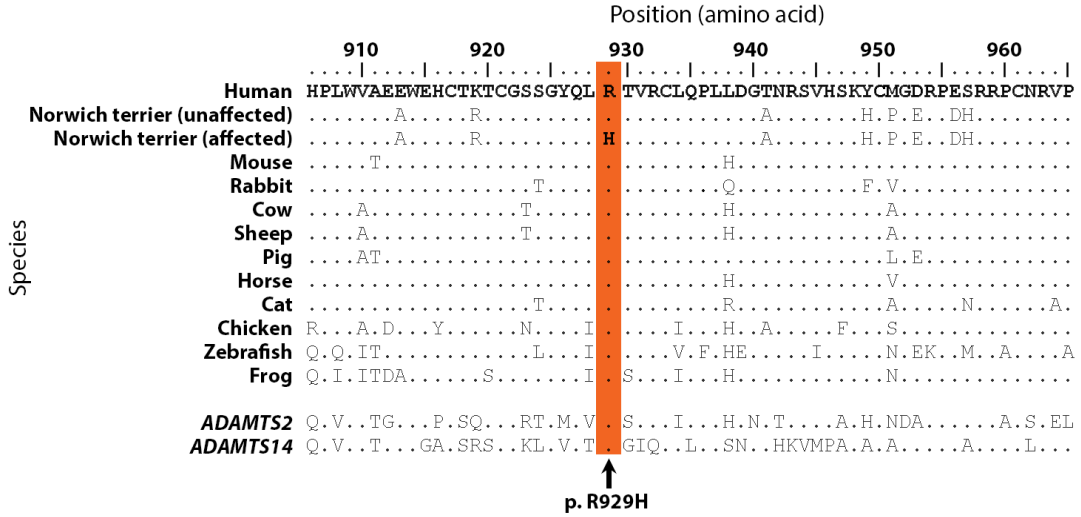


Figure 6.4 - Amino Acid Conservation across Species. Predicted amino acid sequences of the thrombospondin-like domain of vertebrate homologs surrounding the p.R929H missense variant in *ADAMTS3* (black arrow). Sequences are conserved across species at the position of the missense mutation in the Norwich Terrier. Paralogs of human *ADAMTS3*, *ADAMTS2* and *ADAMTS14* are included.

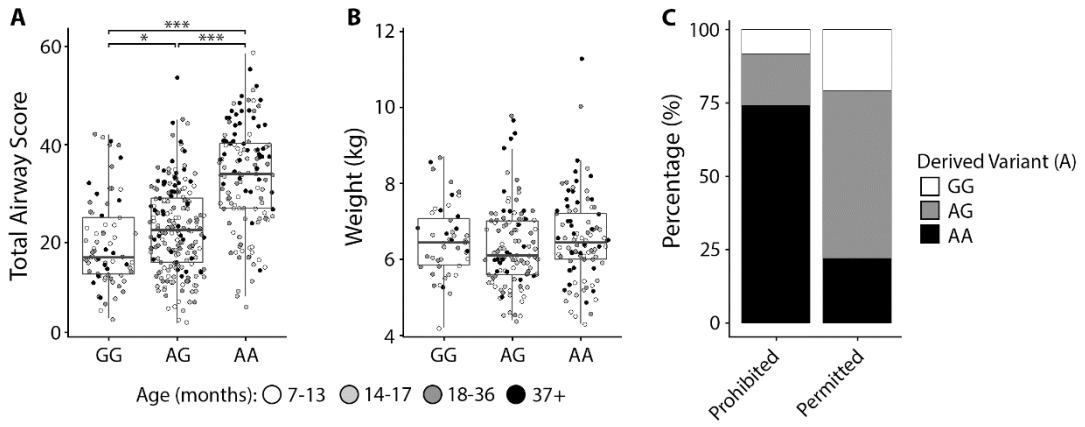


Figure 6.5 - Genotype-Phenotype Correlation of the *ADAMTS3*:c.2786G>A Missense Variant. Distributions of (A) total airway scores and (n = 401) (B) weight (n = 250) for the *ADAMTS3* c.2786G>A allele across screened Norwich Terriers. Individuals are coloured by age quartile at the time of upper airway screening. (C) Under the Swiss Terrier Club guidelines, 78% of the Norwich Terriers screened would be permitted to breed whilst 22% would be prohibited. The *ADAMTS3* c.2786G>A allele within these groups are given as a percentage.

Over 1,300 dogs representing up to 114 diverse breeds including representatives of brachycephalic breeds diagnosed with BOAS were screened for the c.2789G>A variant (Supplementary Table 3). The disease allele frequency (AF) was observed in the Norwich Terrier (AF = 0.57, n = 401), Bulldog (AF = 0.85, n = 41), French Bulldog (AF = 0.12, n = 23), Staffordshire Bull Terrier (AF = 0.125, n = 8), German Spitz (Mittel) (AF = 0.06, n = 8) and Pomeranian (AF = 0.06, n = 8) suggesting the variant may influence BOAS in the French and English Bulldogs. The p.(Arg929His) allele frequency in the human Exome Aggregation Consortium (ExAC) is less than 0.000017 [25].

6.2.7 Discussion

The incidence of UAS amongst the Norwich Terrier population presented a unique opportunity to identify disease modifiers that may be shared across brachycephalic and non-brachycephalic breeds alike. Leveraging laryngoscopic phenotyping, we identified and refined a QTL to an interval that encompasses a single positional candidate gene, *ADAMTS3*. The ADAMTS proteins are a 19-member family of protease enzymes [26,27]. Due to the high sequence similarity, *ADAMTS3* is

categorised in a distinct subgroup of procollagen N-proteinases with ADAMTS2 and -14 [28-30]. *ADAMTS3* is expressed in a variety of tissues including cartilage where it has a substrate specificity for procollagen type II, which it cleaves to stimulate the maturation into collagen II, the major isoform of cartilage [31-35]. In recent years it has become clear that in addition to its role as a procollagen II N-proteinase, *ADAMTS3* also has an important signalling function. *ADAMTS3* proteolytically activates vascular endothelial growth factor-C (VEGF-C), thereby regulating lymphangiogenesis [36,37]. Loss of this signalling function in humans causes Hennekam lymphangiectasia-lymphedema syndrome 3, a condition characterised by lymphedema and distinct facial features including hypertelorism and a flat nasal bridge [38,39]. This oedematous phenotype is reflected in two different *Adamts3* knockout mouse lines which were reported to have severe defects in lymphatic development [40-42]. Both knockout lines resulted in perinatal lethality with Ogino et al., reporting death was due to apparent breathing problems. Interestingly, this line also presented with abnormal rib development and significantly anteroposteriorly shortened skulls [40]. Whilst these studies did not specifically examine the tissues of the upper airways, the oedematous phenotype draws parallels with our observations in affected Norwich Terriers with skeletal changes possibly providing a cryptic selection pressure within the breed. Alternatively, the UAS phenotypes associated with *ADAMTS3* c.2786G>A variant suggest that it predisposes dogs carrying the mutant allele to oedema of the upper airway. This may explain why morphologically diverse breeds (Bulldog and Norwich Terrier), which are predisposed to respiratory conditions, carry the *ADAMTS3* ^{929}His allele. In particular, the oedema hypothesis explains why Norwich Terriers carrying the derived allele are at higher risk of developing UAS whilst not possessing the traditional brachycephalic phenotype which has long been assumed to be the primary predisposition to the disease.

The missense variant reported here (c.2786G>A) could not be separated from a further seventy-nine SNVs and small indels during filtering due to the long-range linkage disequilibrium in the dog [43,44]. None of the additional intronic variants were compelling based on their position of cross-species conservation. Conversely,

arginine 929 in *ADAMTS3* is highly conserved across orthologs and its close paralogs, *ADAMTS2* and *ADAMTS14*, from diverse vertebrates. Arginine 929 is located in the third of four thrombospondin type 1 (TSP1) repeats within *ADAMTS3*. The TSP1 repeats are thought to contribute to substrate binding and interactions with the extracellular matrix [45,46]. Although the variant site is outside the catalytically active metalloproteinase domain of *ADAMTS3*, the canine Arg929His substitution could theoretically alter substrate affinity and/or its modulation of downstream signalling pathways. The exact functional impact of the Arg929His substitution requires further investigation.

Beyond primary lymphedema, associations have been made between *ADAMTS3* and osteoarthritis, heart disease, cancer and schizophrenia in humans [47-50]. A rare coding variant in exon 12 of *ADAMTS3* was also associated with reduced human adult height [51]. Variation in height has also been observed in the dog, albeit with variants in a distant family member, *ADAMTS17*, which were associated with various eye diseases (Rickets et al., in press). We saw no correlation between the *ADAMTS3*:c.2786G>A genotype with body weight or height, for which only a subset of individuals had measurements available (data not shown).

The *ADAMTS3* ^{929}His allele associated with UAS in the Norwich Terrier was also detected in French and English Bulldogs diagnosed with BOAS suggesting *ADAMTS3* may be important for effective respiratory function. Our data might suggest that UAS pathogenesis occurs later in life, however such speculation needs to be tested in a longitudinal study. Longitudinal studies could also assess possible correlations between the risk allele and complications during corrective upper airway surgery, where oedema of the upper airway can predispose to a respiratory crisis, or post-operative prognosis in dogs.

Identification of the *ADAMTS3* ^{929}His allele association with UAS in Norwich Terriers is critically important to help explain the mechanism of a common disease process which is poorly understood, providing a diagnostic marker to identify individuals at increased risk of disease.

6.2.8 Acknowledgements

The authors thank the many Swiss and German breeders and owners of Norwich and Norfolk Terriers, clients of the Hospital for Small Animals (University of Edinburgh) and Royal Veterinary College, all whose beloved pets enabled this study, the Dog Biomedical Variant Database Consortium (Gus Aguirre, Catherine André, Danika Bannasch, Doreen Becker, Brian Davis, Cord Drögemüller, Kari Ekenstedt, Kiterie Faller, Oliver Forman, Steve Friedenberg, Eva Furrow, Urs Giger, Christophe Hitte, Marjo Hytönen, Vidhya Jagannathan, Tosso Leeb, Hannes Lohi, Cathryn Mellersh, Jim Mickelson, Leonardo Murgiano, Anita Oberbauer, Sheila Schmutz, Jeffrey Schoenebeck, Kim Summers, Frank van Steenbeck, Claire Wade) for sharing whole-genome resequencing data, and Mr. Jon Hall for insightful discussions. We also thank Richard Mellanby, Susan Armstrong, Julie Hamilton and Jenni Irving-McGrath for their efforts to spearhead biobanking at the Hospital for Small Animals. Array genotyping services were provided by Edinburgh Genomics. Bioinformatics were conducted on EDDIE3, the University of Edinburgh's computer cluster. ONT reagents were generously donated by Dr. Christine Tait-Burkard. Finally, we would like to thank the Exome Aggregation Consortium and the groups that provided exome variant data for comparison. A full list of contributing groups can be found at http://exac.broadinstitute.org/about.

6.2.9 References

1. Harvey CE (1989) Inherited and Congenital Airway Conditions. J Small Anim Pract 30: 184–187. doi:10.1111/j.1748-5827.1989.tb01531.x.
2. Torrez CV, Hunt GB (2006) Results of surgical correction of abnormalities associated with brachycephalic airway obstruction syndrome in dogs in Australia. J Small Anim Pract 47: 150–154. doi:10.1111/j.1748-5827.2006.00059.x.
3. Fasanella FJ, Shivley JM, Wardlaw JL, Givaruangsawat S (2010) Brachycephalic airway obstructive syndrome in dogs: 90 cases (1991-2008). J Am Vet Med Assoc 237: 1048–1051. doi:10.2460/javma.237.9.1048.

4. Heidenreich D, Gradner G, Kneissl S, Dupré G (2016) Nasopharyngeal Dimensions From Computed Tomography of Pugs and French Bulldogs With Brachycephalic Airway Syndrome. Vet Surg 45: 83–90. doi:10.1111/vsu.12418.

5. Cantatore M, Gobbetti M, Romussi S, Brambilla G, Giudice C, et al. (2012) Medium term endoscopic assessment of the surgical outcome following laryngeal saccule resection in brachycephalic dogs. Vet Rec 170: 518. doi:10.1136/vr.100289.

6. Bernaerts F, Talavera J, Leemans J, Hamaide A, Claeys S, et al. (2008) Description of original endoscopic findings and respiratory functional assessment using barometric whole-body plethysmography in dogs suffering from brachycephalic airway obstruction syndrome. Vet J 183: 95–102. doi:10.1016/j.tvjl.2008.09.009.

7. Ginn JA, Kumar MSA, McKiernan BC, Powers BE (2008) Nasopharyngeal turbinates in brachycephalic dogs and cats. J Am Anim Hosp Assoc 44: 243–249. doi:10.5326/0440243.

8. White RN (2011) Surgical management of laryngeal collapse associated with brachycephalic airway obstruction syndrome in dogs. J Small Anim Pract 53: 44–50. doi:10.1111/j.1748-5827.2011.01156.x.

9. Koch DA, Arnold S, Hubler M, Montavon PM (2003) Brachycephalic Syndrome in Dogs. Compendium on Continuing Education for the Practising Veterinarian 25: 48–55.

10. Liu N-C, Sargan DR, Adams VJ, Ladlow JF (2015) Characterisation of Brachycephalic Obstructive Airway Syndrome in French Bulldogs Using Whole-Body Barometric Plethysmography. PLoS ONE 10: e0130741. doi:10.1371/journal.pone.0130741.

11. Liu N-C, Adams VJ, Kalmar L, Ladlow JF, Sargan DR (2016) Whole-Body Barometric Plethysmography Characterizes Upper Airway Obstruction in 3 Brachycephalic Breeds of Dogs. J Vet Intern Med 30: 853–865. doi:10.1111/jvim.13933.

12. Riecks TW, Birchard SJ, Stephens JA (2007) Surgical correction of brachycephalic syndrome in dogs: 62 cases (1991–2004). J Am Vet Med Assoc 230: 1324. doi:10.2460/javma.230.9.1324.
13. Oechtering GU, Pohl S, Schlueter C, Schuenemann R (2016) A Novel Approach to Brachycephalic Syndrome. 2. Laser-Assisted Turbinectomy (LATE). Vet Surg 45: 173–181. doi:10.1111/vsu.12447.
14. Poncet CM, Dupre GP, Freiche VG, Bouvy BM (2006) Long-term results of upper respiratory syndrome surgery and gastrointestinal tract medical treatment in 51 brachycephalic dogs. 47: 137. doi:10.1111/j.1748-5827.2006.00057.x.
15. Bannasch D, Young A, Myers J, Truvé K, Dickinson P, et al. (2010) Localization of Canine Brachycephaly Using an Across Breed Mapping Approach. PLoS ONE 5: e9632. doi:10.1371/journal.pone.0009632.
16. Schoenebeck JJ, Hutchinson SA, Byers A, Carrington B, Rimbault M et al. (2012) Variation of BMP3 Contributes to Dog Breed Skull Diversity. PLoS Genetics 8: e1002849. doi:10.1371/journal.pgen.1002849.
17. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, et al. (2010) A simple genetic architecture underlies morphological variation in dogs. PLoS Biol 8: e1000451. doi:10.1371/journal.pbio.1000451.
18. Marchant TW, Johnson EJ, McTeir L, Johnson CI, Gow A, et al. (2017) Canine Brachycephaly Is Associated with a Retrotransposon-Mediated Missplicing of SMOC2. Curr Biol 27: 1573–1584.e6. doi:10.1016/j.cub.2017.04.057.
19. Quilez J, Short AD, Martínez V, Kennedy LJ, Ollier W, et al. (2011) A selective sweep of >8 Mb on chromosome 26 in the Boxer genome. BMC Genomics 12: 339. doi:10.1186/1471-2164-12-339.
20. Wiles BM, Llewellyn-Zaidi AM, Evans KM, O'Neill DG, Lewis TW (2017) Large-scale survey to estimate the prevalence of disorders for 192 Kennel Club registered breeds. Canine Genet Epidemiol 4: 8. doi:10.1186/s40575-017-0047-3.

21. Johnson LR, Mayhew PD, Steffey MA, Hunt GB, Carr AH, et al. (2013) Upper airway obstruction in Norwich Terriers: 16 cases. J Vet Intern Med 27: 1409–1415. doi:10.1111/jvim.12206.
22. Dietschi E, Ruchti M, Gaillard C, Stich H, Schawalder P (2010) Oberes Luftweg-Syndrom beim Norwich Terrier. Wissenschaftliche Beilage der Albert-Heim-Stiftung, Zeitschrift der Schweizerischen Kynologischen Gesellschaft SKG, no 3.
23. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature Protocols 4: 1073. doi:10.1038/nprot.2009.86.
24. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010) A method and server for predicting damaging missense mutations. Nat Methods 7: 248–249. doi:10.1038/nmeth0410-248.
25. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, et al. (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536: 285–291. doi:10.1038/nature19057.
26. Apte SS (2009) A disintegrin-like and metalloprotease (reprolysin-type) with thrombospondin type 1 motif (ADAMTS) superfamily: functions and mechanisms. Journal of Biological Chemistry 284: 31493–31497. doi:10.1074/jbc.R109.052340.
27. Bekhouche M, Leduc C, Dupont L, Janssen L, Delolme F, et al. (2016) Determination of the substrate repertoire of ADAMTS2, 3, and 14 significantly broadens their functions and identifies extracellular matrix organization and TGF-β signaling as primary targets. Before we are Born Essentials of Embryology and Birth Defects 30: 1741–1756. doi:10.1096/fj.15-279869.
28. Porter S, Clark IM, Kevorkian L, Edwards DR (2004) The ADAMTS metalloproteinases. Biochem J 386: 15–27. doi:10.1042/BJ20040424.
29. Hurskainen TL, Hirohata S, Seldin MF, Apte SS (1999) ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of zinc metalloproteases. General features and genomic distribution of the ADAM-TS family. Journal of Biological Chemistry 274: 25555–25563. doi:10.1074/jbc.274.36.25555.

30. Colige A, Vandenberghe I, Thiry M, Lambert CA, Van Beeumen J, et al. (2001) Cloning and characterization of ADAMTS-14, a novel ADAMTS displaying high homology with ADAMTS-2 and ADAMTS-3. Journal of Biological Chemistry 277: 5756–5766. doi:10.1074/jbc.M105601200.
31. Fernandes RJ, Hirohata S, Engle JM, Colige A, Cohn DH, et al. (2001) Procollagen II amino propeptide processing by ADAMTS-3. Insights on dermatosparaxis. Journal of Biological Chemistry 276: 31502–31509. doi:10.1074/jbc.M103466200.
32. Le Goff C, Somerville RPT, Kesteloot F, Powell K, Birk DE, et al. (2006) Regulation of procollagen amino-propeptide processing during mouse embryogenesis by specialization of homologous ADAMTS proteases: insights on collagen biosynthesis and dermatosparaxis. Development 133: 1587–1596. doi:10.1242/dev.02308.
33. Olderøy MØ, Lilledahl MB, Beckwith MS, Melvik JE, Reinholt F, et al. (2014) Biochemical and structural characterization of neocartilage formed by mesenchymal stem cells in alginate hydrogels. PLoS ONE 9: e91662. doi:10.1371/journal.pone.0091662.
34. Colige A, Ruggiero F, Vandenberghe I, Dubail J, Kesteloot F, et al. (2005) Domains and maturation processes that regulate the activity of ADAMTS-2, a metalloproteinase cleaving the aminopropeptide of fibrillar procollagens types I-III and V. Journal of Biological Chemistry 280: 34397–34408. doi:10.1074/jbc.M506458200.
35. Nagase T, Ishikawa K, Nakajima D, Ohira M, Seki N, et al. (1997) Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. DNA Res 4: 141–150.
36. Jha SK, Rauniyar K, Leppänen V-M, Brouillard P, Vikkula M, et al. (2017) Efficient activation of the lymphangiogenic growth factor VEGF-C requires the C-terminal domain of VEGF-C and the N-terminal domain of CCBE1. Sci Rep 7: 4916. doi:10.1038/s41598-017-04982-1.

37. Jeltsch M, Jha SK, Tvorogov D, Anisimov A, Leppänen V-M, et al. (2014) CCBE1 enhances lymphangiogenesis via A disintegrin and metalloprotease with thrombospondin motifs-3-mediated vascular endothelial growth factor-C activation. Circulation 129: 1962–1971. doi:10.1161/CIRCULATIONAHA.113.002779.
38. Hennekam RC, Geerdink RA, Hamel BC, Hennekam FA, Kraus P, et al. (1989) Autosomal recessive intestinal lymphangiectasia and lymphedema, with facial anomalies and mental retardation. Am J Med Genet 34: 593–600. doi:10.1002/ajmg.1320340429.
39. Brouillard P, Dupont L, Helaers R, Coulie R, Tiller GE, et al. (2017) Loss of ADAMTS3 activity causes Hennekam lymphangiectasia-lymphedema syndrome 3. Hum Mol Genet 26: 4095–4104. doi:10.1093/hmg/ddx297.
40. Ogino H, Hisanaga A, Kohno T, Kondo Y, Okumura K, et al. (2017) Secreted Metalloproteinase ADAMTS-3 Inactivates Reelin. J Neurosci 37: 3181–3191. doi:10.1523/JNEUROSCI.3632-16.2017.
41. Janssen L, Dupont L, Bekhouche M, Noel A, Voz M, et al. (2015) ADAMTS3 activity is mandatory for embryonic lymphangiogenesis and regulates placental angiogenesis. Angiogenesis 19: 53–65. doi:10.1007/s10456-015-9488-z.
42. Bui HM, Enis D, Robciuc MR, Nurmi HJ, Cohen J, et al. (2016) Proteolytic activation defines distinct lymphangiogenic mechanisms for VEGFC and VEGFD. J Clin Invest 126: 2167–2180. doi:10.1172/JCI83967.
43. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, et al. (2004) Extensive and breed-specific linkage disequilibrium in Canis familiaris. Genome Res 14: 2388–2396. doi:10.1101/gr.3147604.
44. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438: 803–819. doi:10.1038/nature04338.
45. Mead TJ, Apte SS (2018) ADAMTS proteins in human disorders. Matrix Biol. doi:10.1016/j.matbio.2018.06.002.

46. Bekhouche M, Colige A (2015) The procollagen N-proteinases ADAMTS2, 3 and 14 in pathophysiology. Matrix Biol 44-46: 46–53. doi:10.1016/j.matbio.2015.04.001.
47. Kawahara C, Forster T, Chapman K, Carr A, Loughlin J (2005) Genetic association analysis of the IGFBP7, ADAMTS3, and IL8 genes as the potential osteoarthritis susceptibility that maps to chromosome 4q. Ann Rheum Dis 64: 474–476. doi:10.1136/ard.2004.027342.
48. Lee CW, Hwang I, Park C-S, Lee H, Park D-W, et al. (2011) Expression of ADAMTS-2, -3, -13, and -14 in culprit coronary lesions in patients with acute myocardial infarction or stable angina. J Thromb Thrombolysis 33: 362–370. doi:10.1007/s11239-011-0673-7.
49. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, et al. (2014) De novo mutations in schizophrenia implicate synaptic networks. Nature 506: 179–184. doi:10.1038/nature12929.
50. Xu Y, Wang Y, Liu H, Shi Q, Zhu D, et al. (2017) Genetic variants in the metzincin metallopeptidase family genes predict melanoma survival. Mol Carcinog 57: 22–31. doi:10.1002/mc.22716.
51. Marouli E, Graff M, Medina-Gomez C, Lo KS, Wood AR, et al. (2017) Rare and low-frequency coding variants alter human adult height. Nature 542: 186–190. doi:10.1038/nature21039.
52. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, et al. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4: 7. doi:10.1186/s13742-015-0047-8.
53. Delaneau O, Howie B, Cox AJ, Marchini J (2013) Haplotype estimation using sequencing reads. Am J Hum Genet 93: 687–696. doi:10.1016/j.ajhg.2013.09.002.
54. Donnelly P, Marchini J (2009) A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. PLoS Genetics 5: e1000529. doi:10.1371/journal.pgen.1000529.

55. Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nat Genet 44: 821. doi:10.1038/ng.2310.
56. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv q-bio.GN.
57. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, et al. (2014) Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet 46: 912–918. doi:10.1038/ng.3036.
58. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, et al. (2011) Integrative genomics viewer. Nat Biotechnol 29: 24–26. doi:10.1038/nbt.1754.
59. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, et al. (2012) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21. doi:10.1093/bioinformatics/bts635.
60. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.
61. Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, et al. (2012) New and continuing developments at PROSITE. Nucleic Acids Res 41: D344–7. doi:10.1093/nar/gks1067.

6.3 Supporting Information

6.3.1 Methods

Within this submitted manuscript, the experimental materials and methods were streamlined to be in keeping with the style required by the journal. Here is an extended description for procedures and additional results that were not included in the manuscript.

6.3.1.1 Participants

All animal experiments were performed according to local regulations. The dogs recruited at The University of Bern which included the Norwich and Norfolk

Terriers were approved by the "Cantonal Committee for Animal Experiments" (Canton of Bern; permits 22/07, 23/10, and 75/16). Dogs recruited from the University of Edinburgh were approved by the R(D)SVS Veterinary Ethical Review Committee (20 16, University of Edinburgh). All dogs in this study were examined with the consent of their owners.

As part of the ethical approval, this study identified archived records of patients that had visited The Hospital for Small Animals, The University of Edinburgh, and received a BOAS diagnosis but had not been invited to join this study. To identify suitable candidates for recruitment, records were searched in the Tristan Veterinary Practice Management software (v1.8.3.914, Orion Engineering Services). The 'Research' tab of the Tristan software was used to find patients diagnosed with BOAS. Search parameters included 'Dogs' of any breed or sex that had attended the HfSA since January 2008 that are thought to be alive ('Alive only') and had the term 'BOAS' in their appointment notes. A candidate list of dogs from the two searches were provided to Professor Richard Mellanby and Susan Campbell who made contact with owners directly. A PERFORMAgene™ saliva swab with instruction for its use, a consent form, a prepaid return envelope and a short summary of the project aims were sent to owners who were willing to be involved.

6.3.1.2 Bacterial Artificial Chromosome (BAC) Isolation

In order to generate a new reference sequence for the CFA13 critical interval, the critical interval coordinates (chr13:61,166,179-61,579,985) were entered into the University of California Santa Cruz (UCSC) online Genome Browser Gateway (https://genome.ucsc.edu/cgi-bin/hgGateway). Using the "In Other Genomes (Convert)" tool under the "View" menu of the genome browser, the critical interval was converted from the Sep. 2011 (Broad CanFam3.1/canFam3) coordinates to May 2005 (Broad/canFam2) coordinates. Under the browser control parameters, "BAC End Pairs" were displayed in full from the drop down menu for the CanFam2 critical interval (chr13:64,104,713-64,517,949). BACs were selected that would overlap one another and extend beyond the boundaries of the critical interval to ensure that all

sequences within the region could be captured by at least one BAC once they are tiled together. In total, three overlapping BACs (IDs: CH82-2F19, CH82-379O18, CH82-101M10) were selected and ordered from the BACPAC Resource Center based at the Children's Hospital Oakland Research Institute (https://bacpacresources.org/home.htm). Three LB agar stab cultures of DH10B *E.coli* containing each BAC clone were received.

A 25 mg/ml stock of chloramphenicol (Corning, 56-75-7) was prepared by adding 40 mL deionised water to the dried powder provided by Corning. LB agar plates containing 12.5 µg/mL chloramphenicol were prepared by melting 500 mL LB agar in a microwave for 5 minutes and adding 250 µL 25 mg/mL stock chloramphenicol. The melted chloramphenicol LB agar was poured into the base of petri dishes (101R20, Thermoscientific) to cover the bottom of the dish (~20 mL). Once the plates were set, they were inverted to prevent condensation forming on the surface of the agar. Plates were sealed in Parafilm® (P7793-1EA, Sigma-Aldrich) and stored at 4 °C until use. A liquid broth containing a final concentration of 12.5 µg/mL chloramphenicol was prepared.

A pipette tip was used to inoculate the BAC stab cultures on to the chloramphenicol LB plates before the plates were incubated overnight at 37 °C (economy incubator size 2, Gallenkamp). Isolated colonies were picked from the plates using a pipette tip and dropped into a 15 mL Falcon™ tube containing 5 mL liquid LB (12.5 µg/mL chloramphenicol) and incubated in a shaking incubator (Innova 4000, New Brunswick Scientific) at 37 °C for 8 hours at 300 rpm. This starter culture was added to 1 L capped conical flasks containing 200 mL liquid LB (12.5 µg/mL chloramphenicol) and incubated for 16 hours at 37 °C at 300 rpm in a chest incubator (Innova 4330, New Brunswick Scientific). In order to isolate the BAC clones from the cultures were followed the midi-scale preparation protocol for the PhasePrep™ BAC DNA Kit (Sigma-Aldrich, NA0100). The midi-preparation protocol does not perform optimally beyond a cell mass of 1,000. To ensure each culture containing BAC clones did not exceed this, a fluorimeter was used to calculate absorbance.

To operate the Ultrospec fluorimeter (2100, Amersham Biosciences), 5 mL of each liquid culture were added to fluorimeter vials (FB55147, FisherBrand) in addition to a control vial containing the stock liquid LB that hadn't been cultured. The control vial was placed in position 1 of the fluorimeter and samples placed in subsequent slots ensuring that the clear transparent side of the vials were placed in parallel to the light source. The fluorimeter lid was closed and the device was turned on. Absorbance settings were selected. And the OD_{600} readouts were recorded. The cell mass was calculated for each sample by multiplying the OD^{600} absorbance by the total culture volume (200 mL) (Table 6.3).

To isolate the BACs from cultures, each 200 mL midi-culture was split into four 50 mL Falcon™ tubes and pelleted in a centrifuge at 4,000 G for 15 minutes. Each pellet was resuspended in 3 mL RNase A/Resuspension Solution from the PhasePrep™ BAC DNA Kit and transferred to a clean 50 mL Falcon™ tube. Cells were lysed by adding 12 mL Lysis Solution, inverting five times and incubating at room temperature for a maximum of 5 minutes. The cell lysis reaction was neutralized with 12 mL chilled Neutralization Solution and inverting ten times before being incubated on ice for 10 minutes. The cell debris was pelleted by centrifuging at 15,000 G for 30 minutes at 4 °C. The clear BAC-containing supernatant was transferred to a fresh 50 mL Falcon™ tube using large orifice pipette tips (LW4040, Pastette®). The supernatant was mixed with 22 mL room temperature isopropanol and gentle mixed

BAC	OD_{600}	Cell Mass	NanoDrop (ng/mL)	260/ 280	260/ 230	Qubit™ (ng/mL)
BAC1 (CH82-2F19)	1.762	352.4	156.3	1.71	1.31	133
BAC2 (CH82-379O18)	1.659	331.8	190.8	1.72	1.36	96
BAC3 (CH82-101M10)	1.748	349.6	211.8	1.73	1.45	138

Table 6.3 – BAC Isolation. A fluorimeter was used to quantify the absorbance (OD_{600}) of each BAC-containing culture and used to predict their cell mass. Once the BACs were isolated, the NanoDrop and Qubit devices were used to calculate BAC concentrations and infer DNA quality.

by inversion before centrifuging at 15,000 G for 20 minutes at 4 °C. The supernatant was carefully discarded before the pellet was washed for 5 minutes with 10 mL room temperature 70% ethanol and centrifuged at 15,000 G for 5 minutes at 4 °C. The supernatant was removed and the pellet was allowed to air dry for 5 minutes. The pellet was resuspended by gently swirling with 5 mL Elution Solution. Any residual RNA was

degraded by adding 5 µL RNase Cocktail and incubating at 60 °C for 15 minutes. The salt content was adjusted by adding 400 µL Sodium Acetate Buffer Solution and mixing gently. Impurities and endotoxins were removed by mixing with 0.9 mL Endotoxin Removal Solution for 30 seconds before being incubated on ice for 15 minutes. The tube was incubated in a 37 °C water bath for 5 minutes before centrifuging at 4,000 G for 5 minutes at room temperature. The clear upper purified BAC-containing phase was transferred to a fresh 15 mL Falcon™ tube containing 5.4 mL room temperature DNA Precipitation Solution. The sample was mixed thoroughly by inversion before centrifuging at 15,000 G for 30 minutes at 4 °C. The supernatant was removed and the pellet was washed with 5 mL room temperature 70 % ethanol before collecting at the bottom of the tube by centrifuging briefly. The supernatant was removed and the pellet was washed and centrifuged again with 2 mL 70 % ethanol for 2 minutes. All ethanol was removed with a pipette and the pellet was allowed to air dry for 10 minutes. BAC clones were resuspended in 40 µL deionized water and stored at 4 °C until later use.

6.3.1.3 BAC Quantification and Confirmation

A NanoDrop 1000 was used to estimate the concentration of BAC clones in samples and to assess their quality/purity (see Section 2.4.1). The Qubit™ dsDNA broad range quantification kit was used to confirm BAC clone concentrations following the protocol outlined previously (see 2.4.2). Results of the BAC quantification are listed in Table 6.1.

In order to confirm the correct BAC clone was received/isolated, primer pairs were designed to amplify small ~100 bp fragments unique to the 5’ and 3’ ends of

the insert for each BAC clone. Within the CanFam2 critical interval of the UCSC online Genome Browser Gateway, the BAC of interest was selected from the genome browser window. In the “Genomic Alignment of Early Replication Cosmid Ends” section, links were given to sequences of the 5’ and 3’ ends of the BAC inserts. These sequences were copied to the Primer3 online primer design toolkit as previously described (see Section 2.5.4). All parameters were kept as previously described with the exception of the “Product Size Ranges” which was modified to 80-120 bp. Primer pairs were BAC-specific and were predicted to generate amplicons of known sizes between 80 and 120 bp. A standard thermocycler programme was used for all primer pairings and their respective purified BAC preparations (see Section 2.5.2). A 52 °C primer annealing temperature and 30 second extension time was used. When the PCR products were ran on a 2% agarose gel (see Section 2.5.6.2), all primer pairs generated amplicons of the expected for their respective BAC clones indicating the correct clones have been supplied and isolated.

6.3.1.4 BAC Sequencing

The three BAC clones were selected to have overlapping sequences which would allow for the production of a single continuous reference sequence for the CFA13 critical interval. This means the three isolated BAC clones could be pooled into a single sample to sequence on a single run of the Oxford Nanopore MinION. A minimum of 2 µg DNA in 45 µL was required. In order to achieve this, 18.8 µL, 25.8 µL and 18.1 µL of BAC1, BAC2 and BAC3, respectively were pooled in a DNA LoBind Microfuge tube (Eppendorf, 022431048) to give a final quantity of 2.5 µg of each BAC (based on Qubit® quantification, see Table 6.1). This exceeded the total volume of 45 µL so the pooled sample was purified using AMPureXP beads (NG2001, Promega). An equal volume of AMPureXP beads (62.7 µL) was added to the pooled sample which was mixed on a rotator (BMR5010-E, Benchmark Scientific) for 5 minutes at room temperature. The sample was briefly collected using a centrifuge and then placed on a DynaMag™-2 magnetic rack (12321D, ThermoFisher Scientific) for 5 minutes to allow the beads to pellet. The supernatant was discarded and the pellet was washed twice with 200 µL 70% ethanol by pipetting over the pellet whilst still on

the magnet. The pellet was dried for 1 minute and then resuspended in 46 µL nuclease free water. The purified pooled sample was quantified using the Qubit® (see Section 2.4.2) to give a concentration of 112 ng/µL.

The purified pooled BAC sample (45 µL) was repaired by incubating at 20 °C with 8.5 µL nuclease free water, 6.5 µL FFPE Repair Buffer and 2 µL FFPE Repair Mix (NEB M6630, New England Biolabs) for 15 minutes. The DNA was isolated from the reaction mix by adding 62 µL AMPureXP beads and incubating for 5 minutes on a rotator at room temperature. The mixture was briefly pooled by centrifugation and placed on a magnetic rack to pellet the DNA-bound AMPureXP beads. The supernatant was removed and the pellet was washed twice with 200 µL 70% ethanol whilst in the magnet rack before being allowed to dry for 1 minute. The pellet was resuspended in 46 µL nuclease free water. The resuspension was quantified using the Qubit® (see Section 2.4.2) to give a concentration of 80 ng/µL.

To add 3' A-tail ends to the DNA, the 45 µL sample was placed in a strip tube with 7 µL Ultra II End-Prep buffer, 3 µL Ultra II End-Prep enzyme mix (NEB E7595, New England Biolabs) and 5 µL nuclease free water before mixing by inversion and spinning down. The mix was incubated at 20 °C for 30 minutes before 65 °C for 30 minutes. The reaction mixture was transferred back to a DNA LoBind microfuge tube where it was mixed with 60 µL AMPureXP beads and incubated on a rotator for 5 minutes at room temperature. The sample was pooled through centrifugation and pelleted using a magnetic rack. The supernatant was removed and the pellet was washed twice with 200 µL 70% ethanol whilst in the magnetic rack before being allowed to dry for 1 minute. The pellet was resuspended in 31 µL nuclease free water. The resuspension sample was quantified using the Qubit® (see Section 2.4.2) to give a concentration of 59.4 ng/µL.

DNA fragments were ligated by incubating the 30 µL resuspension with 20 µL Adaptor mix and 50 µL NEB Blunt/TA Master mix (NEB M0367, New England BioLad) at room temperature for 20 minutes. Forty µL AMPureXP beads were added and incubated on a rotator for 5 minutes at room temperature. The mixture was pooled

briefly by centrifugation and pelleted on a magnetic rack. The supernatant was discarded and 140 µL ABB beads were added and mixed by flicking. The sample was pooled by centrifugation and pelleted on the magnetic rack. The supernatant was removed and the addition of 140 µL ABB beads was repeated. The beads were resuspended by flicking, collected by brief centrifugation and pelleted on a magnetic rack. The supernatant was removed and the pelleted beads were resuspended in 15 µL ELB and incubated for 10 minutes at room temperature. Finally, the sample is pelleted on the magnet and the eluate was transferred to a new DNA LoBind microfuge tube. The library sample was quantified using the Qubit® (see Section 2.4.2) to give a concentration of 87.8 ng/µL.

To prepare the MinION and Flow Cell (SQK-LSK109, R9.4) for sequencing, the MinION Flow cell was inserted into the MinION device by sliding the end up the supporting clip and pressing down so the electrical port connects. The MinKNOW software was opened and the Flow Cell ID number (SQK-LSK109, R9.4) was entered before clicking “Execute” for the quality control step. Following QC, a total of 897 pores were available for sequencing. Priming buffer was prepared by mixing 480 µL running buffer (RBF) and 520 µL nuclease free water. The priming port of the MinION Flow Cell was opened and an empty P1000 pipette, mounted on a pipetman was held vertically over the priming pore. The volume adjust dial on the pipetman was turned slowly until a few microliters of storage buffer was drawn out of the Flow Cell. The Flow Cell was slowly primed with 800 µL priming buffer so as to ensure that no bubbles were introduced. The Flow Cell was left to incubate with priming buffer for 5 minutes. The SpotON port was opened and 200 µL priming buffer was loaded into the priming pore.

Meanwhile, the library was prepared for loading by mixing 35 µL running buffer (RBF), 25.5 µL Library Loading Beads (LLB), 12 µL library prep and 2.5 µL nuclease free water. A total volume of 75 µL of the library mixture was loaded into the SpotON pore in a continuous dropwise fashion. Both the SpotON and priming ports were closed carefully. The MinION device lid was closed and the sequencing run was activated

(without base-calling) using the MinKNOW software. The Oxford Nanopore MinION device coupled with the MinKNOW software generated a total of 7.59 Gb data in the form of 69,454 raw reads in FAST5 file output. This raw output was base-called used the Albacore (v2.0.1) software prior to constructing contigs with Canu (v1.5) and "polishing" the consensus sequence with Nanopolish (v0.8.4). All three stages used default parameters with each programme.

6.3.2 Results

The age of onset is an important factor to determine to ensure that UAS-susceptible dogs carrying the *ADAMTS3*:c.2786G>A allele are not screened prior to disease onset. All 401 Norwich Terriers in the study were screened once and their age at the time of screening plotted (Figure 6.1). At all ages, homozygous carriers for the derived *ADAMTS3* allele appeared to have an increased severity compared to heterozygous and non-carriers of the allele (Figure 6.1A). Heterozygous carriers appeared to have a non-clinical presentation matching the non-carriers until the age range of 24-29 months. With the exception of 6-11 months of age, the homozygous carriers of the *ADAMTS3* allele have a significantly increased total airway score compared to all other genotypes (Figure 6.1B). A significant difference between the total airway score of heterozygous and non-carriers of the *ADAMTS3* allele only identified in the age range of 30+.

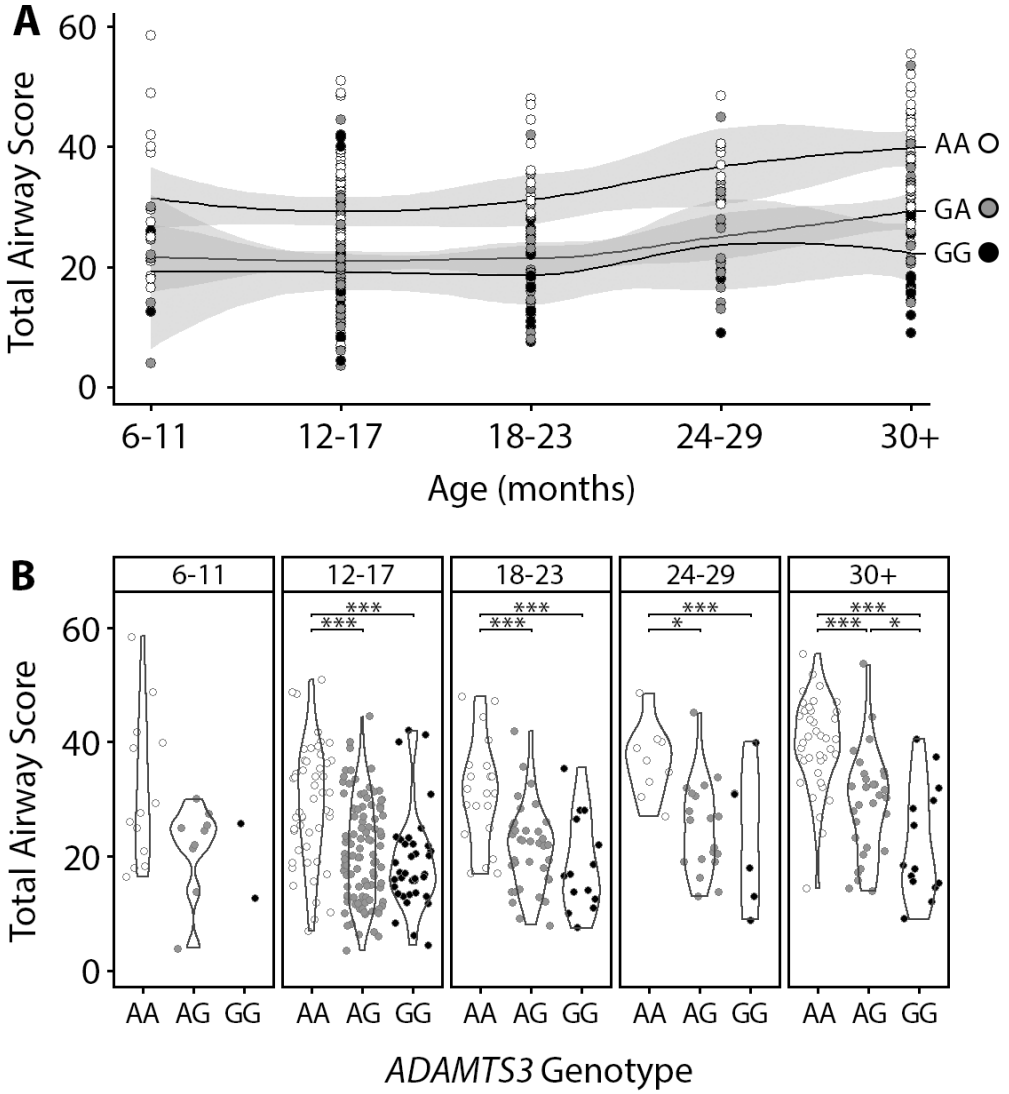


Figure 6.6 – Age of Upper Airway Syndrome Onset. Upper airway scores of all Norwich Terriers plotted at the age of scoping and coloured by *ADAMTS3*:c.2786G>A genotype. (A) Trend lines of disease severity for each genotype and (B) comparison of genotypes at each age point within violin plot overlays. Student's t-test p<0.05 *; <0.01 **; <0.001 *.**

6.4 Discussion

The assessment of the Norwich Terrier upper airway syndrome has proven to give significant insights into not just the suspected cause of the disease in the Norwich Terrier, but also a possible disease modifying risk factor in brachycephalic dogs. Known links between deleterious events in *ADAMTS3* and oedematous disease phenotypes have been shown previously (Brouillard *et al.* 2017) which have made the *ADAMTS3*:c.2786G>A variant discovered here a compelling candidate.

Unfortunately, due to various limitations including the access of specific tissues and the quantity of available data, several follow-up experiments could not be completed.

One follow-up experiment is to comprehensively characterise the skeletal morphologies of the Norfolk and Norwich Terriers and the carriers of the *ADAMTS3*:c.2786G>A variant therein to determine whether the risk allele alters skeletal morphology, as might be expected based on human and mouse data (Marouli *et al.* 2017; Ogino *et al.* 2017). Due to the scarcity of Norwich and Norfolk Terriers receiving CT scans, only a superficial comparison could be made. Importantly, in the context of this manuscript, these Terriers had distinct craniofacial forms and occupied a unique morphospace compared to their brachycephalic counterparts, suggesting brachycephaly per se is not the cause of UAS in the Norwich Terrier. Limited intra-Terrier morphological conclusions could be drawn but suggest that the Norwich and Norfolk Terriers occupy overlapping morphospaces. Variants in *ADAMTS3* have been linked with variation in human height whilst mouse knockouts of the gene have resulted in an anterior-posterior shortening of the skull and abnormal rib development (Marouli *et al.* 2017; Ogino *et al.* 2017). If such morphological variation also exists as a result of the derived variant in *ADAMTS3* in the dog, it may provide a cryptic explanation for its enrichment in the Norwich Terrier over the Norfolk Terrier as breeders could select for a morphological trait. Ideally, to address this question whole body CT scans of a large Norwich and Norfolk Terrier cohort would be collected. Geometric morphometric analyses such as those previously described in this thesis (Chapter 3) could more accurately characterise craniofacial variation between Norwich and Norfolk Terriers as well as between dogs with different *ADAMTS3* genotypes. Furthermore, whole body CT analysis would allow for the assessment of long bone and body axis length variation that have anecdotally been suggested to vary amongst affected and unaffected Norwich Terriers (personal communications with Dr Elisabeth Dietschi, University of Bern).

A second line of follow-up experiments would be to determine the age of onset of the UAS in the Norwich Terrier. Figure 6.6 began to show trends between age at the time of screening and the *ADAMTS3*:c.2786G>A genotype. Individuals homozygous for the derived variant appeared to have an increased severity score over the course of their lifetime whilst the heterozygous state appeared to have a later onset – possibly around 24 months. The data in hand was underpowered for such an experiment due to various reasons. Firstly, to truly determine the age of onset, multiple airway scores would have ideally been taken over the course of an individual's lifetime, however due to the relatively invasive nature of the laryngoscopic screening method, many dogs received just one screen. As a result, if the screen was performed later in life, critical information regarding disease onset was completely missed. A second difficulty was the relatively small size of the study. Whilst 401 participants is a sizeable cohort for many canine disease studies, the division of participants by five age groups and then three genotypes results in too few individuals to meaningfully distil the age of disease onset. Using this current cross sectional study design, only a much larger population cohort, or multiple screens in a lifetime (longitudinal study) would resolve the age of onset – a likely unfeasible feat. Alternatively, a less invasive and easy to repeat scoring system could be acquired such as the use of whole-body barometric plethysmography. This system has been used to score upper airway resistance in brachycephalic breeds and presents no risk to the participants which would permit multiple screens in a lifetime (Liu *et al.* 2015; 2016). Such a scoring system could then be used to possibly identify the functional effects of the *ADAMTS3* variant and track post-operative prognosis.

Critical to the current hypothesis, it would be important to confirm that *ADAMTS3* is expressed in the airway of dogs. This would be expected following the identification of *ADAMTS3* in cartilage of the upper airway of humans (Fernandes *et al.* 2001). In the dog, post-mortems are often carried out at veterinary practices which may present the opportunity to collect appropriate upper airway tissues such as the pharynx, with owner permission. The expression of *ADAMTS3* would not be predicted to vary across genotypes but real-time quantitative PCR would confirm this.

Sections of pharynx could be prepared for immunohistochemistry which would precisely localise *ADAMTS3* and allow for the gross assessment of lymph vessel structure using common markers such as *VEGFR3*, *LYVE1* and *PROX1* which may be expected to vary in carriers of the *ADAMTS3*:c.2786G>A variant (McElroy *et al.* 2008; Vondenhoff *et al.* 2008; Kong *et al.* 2017).

The modulation of collagen maturation in cartilage via procollagen N-proteinase activity was one of the earliest and largely predictable functions of *ADAMTS3* based on its sequence (Nagase *et al.* 1997; Fernandes *et al.* 2001; Colige *et al.* 2005; Le Goff *et al.* 2006; Olderøy *et al.* 2014). As discussed, more recent studies have highlighted its involvement in regulating lymphangiogenesis through VEGF-C modification (Jeltsch *et al.* 2014; Jha *et al.* 2017). A separate study identified a more comprehensive substrate repertoire for *ADAMTS3* by comparing the secretome of two cell lines – one with minimal *ADAMTS3* expression whilst the second was transfected to transiently overexpress *ADAMTS3* (Bekhouche *et al.* 2016). Seventeen peptides were identified as substrate candidates of *ADAMTS3* including peptides of the extracellular matrix (fibronectin and collagen α-chains) and receptors such as TGF-β rector 3 and human leukocyte antigen-A (HLA-A). Enzymatic cleaving of these substrates between the two *ADAMTS3* cell lines were validated through altered banding profiles in Western blots. Similar Western blotting experiments using a cell line transfected with the *ADAMTS3*:c.2786G>A variant could elucidate possible functional ramifications the variant. Furthermore, an overexpressing *ADAMTS3*:c.2786G>A cell line could be used to purify samples of the derived *ADAMTS3* for structural analysis. Liquid chromatography could tease apart subtle changes in the tertiary structure of the protein as a result of the missense variant and begin to explain the altered processing of its substrate repertoire.

Chapter 7: Conclusion and Discussion

At a time when the popularity of brachycephalic dog breeds is growing at an unprecedented level, an understanding of their associated health problems is paramount to address rising welfare concerns. To this end, this thesis endeavoured to determine the genetic risk factors of respiratory disease.

Firstly, a novel computer tomography-based geometric morphometric approach was used to capture variation in canine craniofacial morphology. The intricate relationship of size on shape (allometry) was separated in an optimised manner which saw the use of the neurocranium centroid size and skeletal size-determining genetic loci as regression variables to isolate non-allometric craniofacial shape changes. In doing so, notable intrabreed morphological variation was detected amongst purebred dogs, particularly in breeds such as the Yorkshire Terrier and Labrador Retriever. By necessity, this exploration of the neurocranium centroid size (a proxy for skeletal size), revealed eight loci previously known to be associated with canine size variation. At each of these loci, haplotype mapping defined new critical intervals permitting future studies to investigate their underlying causal variants. In addition to this, the *FGF4* retrogene on CFA18 which was known to arrest development of the limbs was also predicted to dictate variation in the size of the neurocranium.

Importantly, this thesis was able to distil the brachycephaly-associated QTL on CFA1 which had been elusive until now. A 175.6 kb critical interval was defined by individual-level haplotype mapping, which for the first time, could exploit the diversity amongst mixed-breed dogs. *SMOC2* was the only gene residing in the critical interval. In a distinction from previous studies, this thesis aimed to better characterise the potential functional role of the associated gene in craniofacial development. Firstly, *Smoc2* was demonstrated to be expressed in the first pharyngeal arch and frontonasal prominence of both the early chicken and mouse embryos which later give rise to the developing face. The geometric morphometric

assessment of *Smoc2*-null mice were then determined to be craniofacially dysmorphic with similar hallmarks to brachycephalic dogs.

Specifically in the dog, a fragment of a retrotransposon (LINE-1) was identified in intron 8 of *SMOC2* to segregate amongst the brachycephalic dogs of the study cohort. Transcriptomic analysis of the most brachycephalic breeds revealed the integration of this LINE-1 element into *SMOC2* transcripts, thereby introducing a premature a stop codon and truncating potential protein products. Although no proteomic assessment was made, an 80% reduction in *SMOC2* transcript count suggested that the truncated *SMOC2* transcripts were targeted by nonsense-mediated decay. Allele-specific downregulation of *SMOC2* suggested that the LINE-1 variant acted in *cis* and it could not be demonstrated to affect neighbouring genes that flank *SMOC2*. A truncated dominant negative protein product may also be produced but this was not investigated here.

The presentation of respiratory disease in the Norwich Terrier provided an opportunity to parse the confounding influence of the brachycephalic head shape from upper airway disease presentation. A comprehensive characterisation of upper airway syndrome in Norwich Terriers and subsequent GWAS revealed a major QTL on CFA13. Haplotype mapping defined a critical interval to encompass, *ADAMTS3*. In reviewing sequencing data, an error was discovered in the reference assembly within the *ADAMTS3* gene. Long read sequencing technologies corrected this assembly mistake and in doing so revealed 133 bp of missing exon from *ADAMTS3*. Once corrected, a missense coding variant was discovered to segregate with severely affected Norwich Terriers. The scarcity of available tissues restricted downstream functional analysis of the variant, however the derived variant was found to be enriched in the Norwich Terrier and brachycephalic dog breeds including English and French Bulldogs whilst being absent from 109 other breeds.

7.1 Should We Define Canine Brachycephaly?

The term “brachycephaly” was originally used to describe a form of craniosynostosis in humans following the premature fusion of the coronal sutures

(Persing *et al.* 1989). Since then, the term has been adopted by the canine veterinary field to define breed types. Traditionally, the expectation that all individuals of a given breed are morphological uniformed and invariable from one another has resulted in breeds being grouped as "brachycephalic", "mesocephalic" and "dolichocephalic" based purely on assumption. In an attempt to more precisely define such boundaries, several studies sought to quantitatively isolate each subgroup. Evans (Evans 1993) suggested brachycephaly should be defined as having a cranial length to facial length ratio >1.60 whilst Brehm *et al*. (1985) suggested a skull length to skull width ratio <1.44 and Regodón *et al*. (1993) proposed a craniofacial angle between the cranial base and the plane of the hard palate to define morphological types (Brehm *et al.* 1985; Evans 1993; Regodón *et al.* 1993). Most recently, Koch *et al*. (2012) suggested a facial length to cerebrum length of <1.25 for brachycephalic dogs (Koch *et al.* 2012). The lack of consensus across these studies is reflective of the arbitrary values given and the introduction of inherent observer-based biases. For example, the boundary value of 1.25 for brachycephalic dogs by Koch *et al*. (1993) was set to ensure the inclusion of the Norwich Terrier as a brachycephalic breed due to its presentation of upper airway syndrome which was thought to be conformation-induced (Koch *et al.* 2012). Chapter 6 revealed that the Norwich Terrier is in fact morphological distinct from BOAS-diagnosed dogs and the new hypothesis that a susceptibility to oedema rather than a conformation-induced predisposition to UAS is responsible has further disfavoured the use of Koch *et al*. (2012) definition of brachycephaly.

Granted, in Chapter 3 the major source of variability in both the whole head and viscerocranium variability of the dog could be summarised by a brachycephalic-to-dolichocephalic axis. However, along this axis there was no distinction of morphological subgroups, instead morphological variation formed a fluid continuum from one extreme to the other. Importantly, Chapter 3 also highlighted high levels of variation within breeds – more so than would be expected in breeds assumed to be morphologically uniformed. The classically defined "brachycephalic" Dogue de Bordeaux and "mesocephalic" breeds such as the Yorkshire Terrier and Labrador Retriever were amongst some of the most variable breeds. Sources of this

morphological variation is unclear – could there be morphological distinctions between the different functional classes within breeds? For example, is there morphological variability between subgroups of Labrador Retrievers working as 'gun dogs' compared to 'show dogs' or simply 'family pets'? A study of UK Labrador Retrievers (largely non-show dogs) described distinct deviation in dog heights from the UK Kennel Club breed averages but it remains to be seen how the wider roles of the Labrador Retrievers may introduce selection pressures for differing craniofacial morphology (Pugh *et al.* 2015)

The long-standing expectation of morphological uniformity (and genetic homogeneity) of breeds labelled as "brachycephalic" were in part to blame for previous unsuccessful attempts to resolve the brachycephaly-associated CFA1 QTL (Bannasch *et al.* 2010). In reality, phenotypic and genotypic variability within breeds assumed to be brachycephalic prevented the correct definition of a critical interval and isolation of causal variant(s) therein. This thesis eschewed the use of discrete brachycephalic groupings, opting instead to quantify the intra- and inter-breed morphological continuum to rank head shapes. In rejecting discrete subgroupings, and by using a highly specific phenotype (computer tomography), this analysis tolerated natural variation and was able to resolve the CFA1 QTL.

These observations put into question whether a defined boundary for brachycephaly is truly warranted. Collectively, this thesis would be in favour of refraining from the use of misguided boundaries and instead acknowledging both the intra- and inter-breed fluid variability. Since "brachycephaly" represents the extreme of morphological variation, perhaps an appropriate morphological description of dogs at this extreme would be to refer to them as the "most brachycephalic" dogs rather than simply "brachycephalic".

7.2 The Reign of the Structural Variant

Ever since the publication of the first canine draft assembly 15 years ago and its subsequent updates, structural variants such as LINEs and SINEs have been hypothesised to contribute to the extreme variability across the canine population

(Kirkness *et al.* 2003; Lindblad-Toh *et al.* 2005; Wang and Kirkness 2005). The high abundance of the elements which together make up to ~30% of the canine genome were believed to be rapidly enriched during domestication. Despite this, relatively few LINEs and SINEs have been identified as causing canine traits and diseases – until recently. Particularly within the last five years, more attention has been drawn to the family of retrotransposons which have been shown to cause a variety of traits from cleft palates to white spotting on coats (Brooks *et al.* 2004; Credille *et al.* 2009; Smith *et al.* 2010; Dreger and Schmutz 2011; van Steenbeek *et al.* 2013; Wolf *et al.* 2014; Downs and Mellersh 2014; Körberg *et al.* 2014; Wiedmer *et al.* 2015; Marchant *et al.* 2017; Mauri *et al.* 2017; Everson *et al.* 2017). The previous expense of whole genome sequencing made it difficult to detect the often intronic localisation of these large-spanning elements. Even today, the reliance on short-read data and the lack of a standardised pipeline for detecting structural variants comparable to the GATK best practice guidelines (DePristo *et al.* 2011), have resorted in geneticists visually inspecting genomes for hallmarks of retrotransposons (Downs and Mellersh 2014; Mauri *et al.* 2017; Everson *et al.* 2017). Ongoing developments in long-read sequencing platforms will likely aid in the discovery of large structural variants in the future.

The characterisation of the brachycephaly associated LINE-1 element within intron 8 of *SMOC2* in Chapter 5 adds to the mounting literature on the detrimental effects of retrotransposons. Through the introduction of alternative splice sites situated within the introduced LINE-1, varying portions of LINE-1 sequence is incorporated into *SMOC2* transcripts after exon 8 on the canonical 13 exon gene. As a result, novel LINE-1 derived stop codons are predicted to prematurely terminate translation. It is unclear whether a protein product is formed however the absence of the exons ten through thirteen, which encode the protein's extracellular calcium-binding (EC) domain, is likely to render the truncated protein non-functional. The functional dependency of this domain has been proven in cell lines that lose migratory and differentiation regulatory properties once the EC domain of *SMOC2* was removed (Maier *et al.* 2008; Peeters *et al.* 2018). Furthermore, the ability of BM-

40 members to bind substrates such as the collagens is calcium-dependent, highlighting the importance of the EC domain (Mayer *et al.* 1991; Sasaki *et al.* 1997; 1998). Despite these predictions, it remains plausible that the truncated protein could act in a dominant-negative fashion, for example by generating protein aggregates and inducing toxicity (Veitia 2007).

Regardless of whether a truncated protein product exists, the introduction of the LINE-1 fragment in a homozygous state causes an 80% loss of full-length *SMOC2* transcripts. Comparably, intronic LINE-1 insertions creating premature stop codons in the distal-less homeobox 6 (*DLX6*) and transglutaminase 1 (*TGM1*) genes reduce full transcript counts by ~75% and ~85% of their respective genes (Credille *et al.* 2009; Wolf *et al.* 2014). These LINE-1 insertions in *DLX6* and *TGM1* are linked with cases of cleft palates and ichthyosis in dogs, respectively. Across the *SMOC2*, *DLX6* and *TGM1* examples, the LINE-1 fragments varied in length (6.2 kb, 2.1 kb, 2.0 kb) and were inserted after exons 8 of 13, 9 of 14 and 2 of 3, respectively. An additional intronic LINE-1 insertion of 2.0 kb after exon 5 of 8 in the factor IX (F9) gene was linked with mild haemophilia B in dogs however no expression analysis was performed (Brooks *et al.* 2004). It remains unclear to what extent the size of the LINE-1, which intron it resides in, the proximity to exons and its orientation may have on the host gene's relative expression, which warrants further evaluation in future studies. Nevertheless, the effectiveness of LINE-1's to cause morphological and disease traits is evident. As whole genome sequencing becomes cheaper, supplemented by advances in long-read technologies which can span these structural variants, the number of traits dictated by retrotransposons is likely to escalate further in the coming years.

7.3 A New Perspective on Brachycephalic Obstructive Airway Syndrome

The almost exclusive presentation of BOAS amongst the most brachycephalic short-faced dogs has long supported the broadly held belief that head conformation is responsible for the condition. The rationale being that the soft tissues of the shortened muzzle do not shrink proportionally with their bony framework and so

obstruct the airway causing the characteristic symptoms (Ladlow *et al.* 2018). The influence of weight and age on the condition have proven inconclusive and yet a heterogeneous disease presentation remains across and within breeds (2006; Torrez and Hunt 2006; Fasanella *et al.* 2010). In chapter 6, the discovery of the *ADAMTS3*:c.2786G>A variant in Norwich Terriers and its link to an oedematous upper airway phenotype has given new insights into the aetiology of respiratory disease. In addition to their compromising head shape, one in four French Bulldogs carry the *ADAMTS3*:c.2786G>A disease allele whilst in the English Bulldog this allele is approaching fixation. These observations challenge old perspectives of disease causation of BOAS and present a new hypothesis:

Oedema of the airway caused by the disruption of *ADAMTS3* predisposes its host to upper airway obstruction with or without the brachycephalic skull conformation.

The analysis of CT data suggests that the *ADAMTS3*:c.2786G>A allele does not affect head shape. In finding a novel, non-conformational modifier of respiratory disease in dogs, it has potentially wide-reaching ramifications in the veterinary and breeding communities. By necessity, this thesis developed an easy PCR-based method to genotyping the variant which can readily be replicated elsewhere. For the Norwich Terrier community, a screening programme for upper airway syndrome already exists – albeit using a physical examination of the airway. The willingness of the Swiss Terrier Club to use such a programme provides a framework upon which a genetic screen can be rolled out in place of/alongside the current physical examination. Hopefully the ease and reduced risk of a genetic screen would continue to reduce UAS severity amongst Norwich Terriers as already started by the airway screening programme.

Currently, there are no existing large-scale programmes designed to tackle BOAS amongst the brachycephalic dogs. The importance of the physical trait conformation set down by the breed standards and avidly adhered to by breeders has limited the management of BOAS. For the French Bulldog at least, the discovery

of the *ADAMTS3* variant at manageable frequencies (~0.25) provides the opportunity to selectively breed to reduce transmission of the risk variant to future generations of French Bulldog population with minimal, or potentially no changes to morphological traits. The fine-level impact of the *ADMATS3*:c.2786G>A allele on morphological variation is yet to be determined but it is expected to be limited due to high morphological similarities between the Norwich Terrier (carriers) and Norfolk Terrier (non-carriers).

BOAS management in the English Bulldog may prove more difficult since the risk variant is nearly fixed in this population; of the twenty-four English Bulldogs tested, nineteen (79.2%) were homozygous carriers of the *ADAMTS3*:c2786G>A allele, three (12.5%) were heterozygous carriers whilst just 2 (8.3%) were free of the mutation. The enrichment of the deleterious *ADAMTS3* variant is perhaps an uncomfortable testament to the high genetic homogeneity amongst English Bulldogs as the result of genetic bottlenecks and a small founder population (Pedersen *et al.* 2016). It is estimated that there are just four paternal haplotypes in the breed with one haplotype occurring in 93% of dogs. Selective breeding to purge the risk allele from future English Bulldog generations risks exacerbating inbreeding further. To this end, the feasible resolution of the problematic *ADAMTS3* variant, alongside managing a wide range of other detrimental congenital health defects whilst maintaining desirable traits from within the breed is questionable (Pedersen *et al.* 2016). The combination of extreme brachycephaly and the *ADAMTS3* variant explains the greater severity of pathology observed in Bulldogs with BOAS .

7.4 Future Directions

A persistent challenge for the field of canine genetics and genomics is the investigation of the biological relevance of positional candidate genes and their putatively causal variants. This thesis endeavoured to address such issues for the brachycephaly-associated LINE-1 insertion in *SMOC2*. Follow-up experiments included *in silico* modelling of protein function, localisation of embryonic *SMOC2* expression, comparison of transcriptomic profiles across *SMOC2* genotypes in dogs,

comparison of sequence homology of predicted transcripts across species and the assessment of craniofacial structure in *Smoc2*-null mice. Collectively, these experiments strongly inferred the role of *SMOC2* and its associated LINE-1 variant with canine brachycephaly. Unfortunately, causal *proof* could not be ascertained. Ultimately, proof would only be achieved through the introduction of the LINE-1 element into a non-brachycephalic dog through the use of genome editing tools and subsequent craniofacial analysis. The myostatin (*MSN1*) gene was edited in the dog using a clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) 9 system to create a double-muscled Dachshund (Zou *et al.* 2015). In doing so, the viability of the system in the dog was proven however the ethical and moral responsibility of such experiments in the dog is questionable.

A feasible alternative to aid the proof of causality would be to make comparable edits to the mouse genome. The CRISPR/Cas 9 system has revolutionised genome editing, particularly in the mouse where it was first described (alongside human cells) (Ran *et al.* 2013). The relative ease, quicker turnaround of edited offspring and higher efficiency of the system means multiple mouse lines could be generated in parallel to explore the role of the region on phenotypic output (Burgio 2018). The predicted effect of the dog LINE-1 in *SMOC2* is to truncate the protein after exon 8. Importantly, ancestral versions of *SMOC2* are still produced (see Chapter 5) and so ablating exons 9 to 13 in the mouse would not be a reflective model of the dog. Instead, the LINE-1 insertion in dogs is ~6.3 Kb in size which could be inserted into the homologous region of the mouse intron 8 using a single guide RNA and a vector plasmid containing the LINE-1 to induce homology-directed repair. Alongside this, a number of additional mouse lines would help elucidate the function of the locus. Currently, it is unclear what, if at all, the intronic sequence in the immediate vicinity of the LINE-1 may have on its role in splicing *SMOC2*. To this end, appropriate further mouse modifications would be to swap the murine intron 8 for the dog intron 8 – a “caninisation” of the mouse gene. Such an experiment would need to be carried out for both the derived (containing the LINE-1) and ancestral (no LINE-1) versions of the dog intron 8 which could be achieved by a pair of guide RNAs

either side of the mouse target intron. The ancestral "caninised" mouse would act as a control to the derived "caninisation" of the gene whilst an additional wildtype mouse on the same background as the edited mice would act as a true control to any edits.

Once the mouse models were generated, an array of experiments would be available. The primary interest would be in the end-stage phenotype i.e. do the mice harbouring a LINE-1 edit have dysmorphic skulls? A micro-CT scan of the mature skull (6 month) would permit a geometric morphometric analysis of craniofacial variation similar to that used here for dogs (Kristensen *et al.* 2009; Campbell and Sophocleous 2014). The maintenance of mouse lines would also grant access to embryonic tissue throughout development. *Smoc2* is predicted to regulate cell migration, the maturation of pre-osteoblast and the mineralisation of matrix (Rocnik *et al.* 2006; Liu *et al.* 2007; Maier *et al.* 2008; Peeters *et al.* 2018). Neural crest cells (NCC) represent a major migratory cell population which translocate from the dorsally positioned neural tube to the frontonasal prominence and pharyngeal arches (Nichols 1981; Erickson and Weston 1983; Calloni *et al.* 2009). All bone of the viscerocranium, the most prominently affected anatomical region to be altered in brachycephalic dogs, is entirely derived from NCCs (Chai *et al.* 2000; Jiang *et al.* 2002). Furthermore, the mobilisation of NCC are particularly sensitive to subtle variation in the surrounding extracellular matrix – something that may be induced by alterations in the quantity and possibly function of *Smoc2* (Perris 1997; Henderson and Copp 1998; Perris and Perissinotto 2000). Therefore, an intriguing investigation would be to monitor NCC behaviour over the different mouse lines. Immunohistochemistry (IHC) staining of migrating NCCs using markers such as *Pax3* or *Foxd3* (Epstein *et al.* 1991; Dottori *et al.* 2001) could infer comparative migration at different embryonic time points. Also, the loss of the extracellular calcium-binding domain of *Smoc2* has been shown *in vitro* to promote premature differentiation into osteoblasts and mineralisation of the surrounding matrix (Peeters *et al.* 2018). This could be confirmed by IHC staining of whole-mount sections with the osteoblast marker, osteonectin (Schulz *et al.* 1988) to

infer pre-osteoblast differentiation and cartilage (Alizarin Blue) and mineralisation (Alizarin Red) markers to detect endochondral ossification (Rigueur and Lyons 2014).

SMOC2 is considered to be an important signal mediator by acting as an intermediary protein between various growth factors, the extracellular matrix and cell surface receptors (Rocnik *et al.* 2006; Liu *et al.* 2007; Maier *et al.* 2008). To this end, *SMOC2* would be expected to influence the proteomic makeup of the tissues in which it is expressed. To investigate the proteomic networks around *Smoc2*, the pharyngeal arches, which are known to express *Smoc2*, could be excised from the previously generated mouse lines and prepared for proteomic analysis, possible with a labelled quantitative mass spectrometry approach such as iTRAQ (Wiese *et al.* 2006). Furthermore, through the selected RT-qPCR in zebrafish, key members of the *Bmp* signalling network were shown to be downregulated in *Smoc2* mutants (Mommaerts *et al.* 2014). A full transcriptomic analysis of *Smoc2* variants has not been performed which again, could be executed in the pharyngeal arches of the generated mouse lines.

The identification of the *ADAMTS3*:p.Arg929His allele as a disease modifying factor in conjunction with the confounding craniofacial shape influenced by the newly discovered *SMOC2* LINE-1 insertion and previously known *BMP3*:p.Phe452Leu allele warrant the question; can these variants accurately predict BOAS severity? Or importantly in a veterinary setting, can these variants predict the likelihood of complications during surgery and a patient's post-surgical success? To address this question, longitudinal studies that utilise whole-body barometric plethysmography could be used to assess upper airway function of brachycephalic breeds before and throughout disease onset over the course of their life and treatment. In parallel, non-brachycephalic healthy breeds could be assessed as a comparative control which would not normally be subject to invasive airway assessments such as laryngoscopy. The objectivity of the barometric plethysmography coupled with its growing popularity amongst veterinarians means a UK-wide recruitment of patients could be achieved to increase the cohort size and overall power of the study. Genotypes for

the *ADAMTS3*, *SMOC2* and *BMP3* variants would be used as variables for upper airway function in a multivariate analysis. The results of this study would indicate the single, or combination of genotypes that have the greatest influence on BOAS severity, surgical complications and post-surgical prognosis. Therefore, veterinarians could screen for *ADAMTS3*, *SMOC2* and *BMP3* as predicative indicators of surgical success in their pre-surgical assessment of patients or highlight cases that are likely to warrant more intense post-operative care.

Beyond the scope of the veterinary community, the discovery of variants in *SMOC2* and *ADAMTS3* may translate to the field of human medicine. *SMOC2* is already known to at least partly have a role in craniofacial development in humans. Specifically, an isolated family with severe dental abnormalities including microdontia and oligodontia were attributed to missense mutations within *SMOC2* (Bloch-Zupan *et al.* 2011). It is unclear whether more substantial variants within the *SMOC2* gene could have more global effects on craniofacial morphology. However, with 1 in 2,100 new born children having craniofacial abnormalities and just 20% of them receiving a genetic diagnosis (Lajeunie *et al.* 1995; Persing 2008; Johnson and Wilkie 2011), *SMOC2* could offer a new gene to screen in cases where the causal genetic variant is unknown.

Also, like dogs, humans suffer from obstructive respiratory diseases thought to be linked with variation in craniofacial shape. Namely, obstructive sleep apnoea (OSA) is a chronic respiratory condition caused by the repetitive partial or complete collapse of the upper airway during sleep (Lattimore *et al.* 2003). Sufferers commonly experience fragmented sleep, oxygen desaturation and are at risk of cardiovascular disease and other forms of mortality (Sin *et al.* 1999; Lattimore *et al.* 2003). It is predicted that up to 3.9% and 1.2% of the adult US male and female populations are affected by OSA, respectively (Bixler *et al.* 1998; Sin *et al.* 1999; Bixler *et al.* 2001). A single craniometric measurement, facial depth at the height of the palate can account for up to 15% of OSA severity (apnoea-hypopnea index), making it the second greatest explanatory variable after obesity (Dempsey *et al.* 2002). Separately, face

width, eye width and mandibular length can predict OSA with a sensitivity of 86% (Lee *et al.* 2009). These craniofacial morphological variations are consistent with brachycephalic head formation. An observation supported by Capistrano et al., 2015 who identified a significant increase in apnoea-hypopnea index to be associated with the brachycephalic facial morphology compared to dolicho- and mesocephalic (Capistrano *et al.* 2015). The co-segregation of *SMOC2* variants with respiratory distress in dogs could suggest variability in human *SMOC2* could predispose humans to OSA. Furthermore, obstructive respiratory conditions in both dogs and humans are likely to have other non-conformation-related factors. Like the proposed oedematous airway hypothesis in upper airway syndrome of the Norwich Terrier, humans also experience severe respiratory obstruction caused by the predisposition to oedema of the upper airway. Laryngeal lymphedema and hereditary lymphedema are both debilitating conditions which amongst other symptoms, can cause severe breathing difficulties (López-Lera *et al.* 2011; Deng *et al.* 2012; Smith *et al.* 2014). To date, the aetiology of these conditions is restricted to the association of variants in the serpin family G member 1 (*C1NH*) gene with hereditary lymphedema (López-Lera *et al.* 2011). The identification of *ADAMTS3* variants in canine cases of oedematous obstruction of the airway could warrant a screen for similar variants humans that suffer from these poorly understood conditions.

As discussed here and previously, further experiments are warranted to dissect the biological effects of *SMOC2*/*ADAMTS3* variants within their respective tissues and to interrogate the likely complex interplay in their predisposition of respiratory morbidities. Follow-up studies also have the capacity to uncover insights into human health however the fundamental implications of this thesis remain grounded in the field of veterinary medicine. The discovery of a predisposition to oedema in Upper Airway Syndrome and its likely contribution to specific breeds with Brachycephalic Obstructive Airway Syndrome have substantial ramifications for the management of the debilitating respiratory diseases across the Norwich Terrier and brachycephalic dog breeds alike. Particularly, the discovery of a potential role for a non-conformation-related risk factor to Brachycephalic Obstructive Airway Syndrome

challenges current perceptions of this high prevalence condition. Hopefully, though the careful and considered use of this data, the welfare of breeds susceptible to respiratory diseases will be improved.

Reference List

Abzhanov A., Rodda S. J., McMahon A. P., Tabin C. J., 2007 Regulation of skeletogenic differentiation in cranial dermal bone. Development **134**: 3133–3144.

Adhikari K., Fuentes-Guajardo M., Quinto-Sánchez M., Mendoza-Revilla J., Chacón-Duque J. C., Acuña-Alonzo V., Jaramillo C., Arias W., Lozano R. B., Pérez G. M., Gómez-Valdés J., Villamil-Ramírez H., Hunemeier T., Ramallo V., de Cerqueira C. C. S., Hurtado M., Villegas V., Granja V., Gallo C., Poletti G., Schuler-Faccini L., Salzano F. M., Bortolini M.-C., Canizales-Quinteros S., Cheeseman M., Rosique J., Bedoya G., Rothhammer F., Headon D., González-José R., Balding D., Ruiz-Linares A., 2016 A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. Nat Commun **7**: 11616.

Alazami A. M., Shaheen R., Alzahrani F., Snape K., Saggar A., Brinkmann B., Bavi P., Al-Gazali L. I., Alkuraya F. S., 2009 FREM1 mutations cause bifid nose, renal agenesis, and anorectal malformations syndrome. Am J Hum Genet **85**: 414–418.

Alfawaz S., Fong F., Plagnol V., Wong F. S. L., Fearne J., Kelsell D. P., 2013 Recessive oligodontia linked to a homozygous loss-of-function mutation in the SMOC2 gene. Arch Oral Biol **58**: 462–6.

Anderson P. J., Hall C. M., Evans R. D., Jones B. M., Hayward R. D., 1997 Hand anomalies in Crouzon syndrome. Skeletal Radiol **26**: 113–115.

Apert E., 1906 De l'acrocéphalosyndactylie. Bulletins et Mémoires de la Société Medicale des Hôpitaux de Paris **23**: 1310.

Arnason U., Adegoke J. A., Bodin K., Born E. W., Esa Y. B., Gullberg A., Nilsson M., Short R. V., Xu X., Janke A., 2002 Mammalian mitogenomic relationships and the root of the eutherian tree. Proc Natl Acad Sci USA **99**: 8151–8156.

Asher L., Diesel G., Summers J. F., McGreevy P. D., Collins L. M., 2009 Inherited defects in pedigree dogs. Part 1: Disorders related to breed standards. Vet J **182**: 402.

Ayling R. M., Ross R., Towner P., Laue Von S., Finidori J., Moutoussamy S., Buchanan C. R., Clayton P. E., Norman M. R., 1997 A dominant-negative mutation of the growth hormone receptor causes familial short stature. Nat Genet **16**: 13–14.

Baik I., Seo H. S., Yoon D., Kim S. H., Shin C., 2015 Associations of Sleep Apnea, NRG1 Polymorphisms, Alcohol Consumption, and Cerebral White Matter Hyperintensities: Analysis with Genome-Wide Association Data. J Appl Physiol **38**: 1137–1143.

Bannasch D., Young A., Myers J., Truvé K., Dickinson P., Gregg J., Davis R., Bongcam-Rudloff E., Lindblad-Toh K., Pedersen N., 2010 Localization of Canine Brachycephaly Using an Across Breed Mapping Approach. PLoS ONE **5**: e9632.

Bannink N., Maliepaard M., Raat H., Joosten K. F. M., Mathijssen I. M. J., 2011 Obstructive sleep apnea-specific quality of life and behavioral problems in children with syndromic craniosynostosis. J Dev Behav Pediatr **32**: 233–238.

Bannink N., Nout E., Wolvius E. B., Hoeve H. L. J., Joosten K. F. M., Mathijssen I. M. J., 2010 Obstructive sleep apnea in children with syndromic craniosynostosis: long-term respiratory outcome of midface advancement. Int J Oral Maxillofac Surg **39**: 115–121.

Beaty T. H., Murray J. C., Marazita M. L., Munger R. G., Ruczinski I., Hetmanski J. B., Liang K. Y., Wu T., Murray T., Fallin M. D., Redett R. A., Raymond G., Schwender H., Jin S.-C., Cooper M. E., Dunnwald M., Mansilla M. A., Leslie E., Bullard S., Lidral A. C., Moreno L. M., Menezes R., Vieira A. R., Petrin A., Wilcox A. J., Lie R. T., Jabs E. W., Wu-Chou Y. H., Chen P. K., Wang H., Ye X., Huang S., Yeow V., Chong S. S., Jee S. H., Shi B., Christensen K., Melbye M., Doheny K. F., Pugh E. W., Ling H., Castilla E. E., Czeizel A. E., Ma L., Field L. L., Brody L., Pangilinan F., Mills J. L., Molloy A. M., Kirke P. N., Scott J. M., Scott J. M., Arcos-Burgos M., Scott A. F., 2010 A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nat Genet **42**: 525–529.

Beck C. R., Garcia-Perez J. L., Badge R. M., Moran J. V., 2011 LINE-1 elements in structural variation and disease. Annu Rev Genomics Hum Genet **12**: 187–215.

Beederman M., Farina E. M., Reid R. R., 2014 Molecular basis of cranial suture biology and disease: Osteoblastic and osteoclastic perspectives. Genes Dis **1**: 120–125.

Bekhouche M., Leduc C., Dupont L., Janssen L., Delolme F., Goff S. V.-L., Smargiasso N., Baiwir D., Mazzucchelli G., Zanella-Cleon I., Dubail J., De Pauw E., Nusgens B., Hulmes D. J. S., Moali C., Colige A., 2016 Determination of the substrate repertoire of ADAMTS2, 3, and 14 significantly broadens their functions and identifies extracellular matrix organization and TGF-β signaling as primary targets. (KL Moore and TVN Persaud, Eds.). Before we are Born Essentials of Embryology and Birth Defects **30**: 1741–1756.

Benecke N., 1987 Studies on early dog remains from Northern Europe. Journal of Archaeological Science **14**: 31.

Bernaerts F., Talavera J., Leemans J., Hamaide A., Claeys S., Kirschvink N., Clercx C., 2008 Description of original endoscopic findings and respiratory functional assessment using barometric whole-body plethysmography in dogs suffering from brachycephalic airway obstruction syndrome. Vet J **183**: 95–102.

Berryere T. G., Kerns J. A., Barsh G. S., Schmutz S. M., 2005 Association of an Agouti allele with fawn or sable coat color in domestic dogs. Mamm Genome **16**: 262–272.

Bi W., Deng J. M., Zhang Z., Behringer R. R., de Crombrugghe B., 1999 Sox9 is required for cartilage formation. Nat Genet **22**: 85–89.

Bialek P., Kern B., Yang X., Schrock M., Sosic D., Hong N., Wu H., Yu K., Ornitz D. M., Olson E. N., Justice M. J., Karsenty G., 2004 A twist code determines the onset of osteoblast differentiation. Dev Cell **6**: 423–435.

Bixler E. O., Vgontzas A. N., Have Ten T., Tyson K., Kales A., 1998 Effects of age on sleep apnea in men: I. Prevalence and severity. Am J Respir Crit Care Med **157**: 144–148.

Bixler E. O., Vgontzas A. N., Lin H. M., Have Ten T., Rein J., Vela-Bueno A., Kales A., 2001 Prevalence of sleep-disordered breathing in women: effects of gender. Am J Respir Crit Care Med **163**: 608–613.

Björnerfeldt S., Webster M. T., Vilà C., 2006 Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Res **16**: 990–994.

Bloch-Zupan A., Jamet X., Etard C., Laugel V., Muller J., Geoffroy V., Strauss J.-P., Pelletier V., Marion V., Poch O., Strahle U., Stoetzel C., Dollfus H., 2011 Homozygosity mapping and candidate prioritization identify mutations, missed by whole-exome sequencing, in SMOC2, causing major dental developmental defects. Am J Hum Genet **89**: 773–81.

Bochukova E. G., Roscioli T., Hedges D. J., Taylor I. B., Johnson D., David D. J., Deininger P. L., Wilkie A. O. M., 2008 Rare mutations of FGFR2 causing apert syndrome: identification of the first partial gene deletion, and an Alu element insertion from a new subfamily. Hum Mutat **30**: 204–211.

Bonilla-Claudio M., Wang J., Bai Y., Klysik E., Selever J., Martin J. F., 2012 Bmp signaling regulates a dose-dependent transcriptional program to control facial skeletal development. Development **139**: 709–719.

Bookstein F. L., 1992 Distance Measures. In: *Morphometric Tools for Landmark Data: Geometry and Biology*, Cambridge University Press, pp. 88–124.

Boyko A. R., Boyko R. H., Boyko C. M., Parker H. G., Castelhano M., Corey L., Degenhardt J. D., Auton A., Hedimbi M., Kityo R., Ostrander E. A., Schoenebeck J., Todhunter R. J., Jones P., Bustamante C. D., 2009 Complex population structure in African village dogs and its implications for inferring dog domestication history. Proc Natl Acad Sci USA **106**: 13903–13908.

Boyko A. R., Quignon P., Li L., Schoenebeck J. J., Degenhardt J. D., Lohmueller K. E., Zhao K., Brisbin A., Parker H. G., vonHoldt B. M., Cargill M., Auton A., Reynolds A., Elkahloun A. G., Castelhano M., Mosher D. S., Sutter N. B., Johnson G. S., Novembre J., Hubisz M. J., Siepel A., Wayne R. K., Bustamante C. D., Ostrander E. A., 2010 A simple genetic architecture underlies morphological variation in dogs. PLoS Biol **8**: e1000451.

Breathnach R., Benoist C., O'Hare K., Gannon F., Chambon P., 1978, Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-

intron boundaries. Proceedings of the National Academy of Sciences of the United States of America. **75**: 4853-4857.

Breen M., Modiano J. F., 2008 Evolutionarily conserved cytogenetic changes in hematological malignancies of dogs and humans--man and his best friend share more than companionship. Chromosome Res **16**: 145–154.

Brehm H., Loeffler K., Komeyli H., 1985 [Skull forms in dogs]. Anat Histol Embryol **14**: 324–331.

Brekken R. A., Sage E. H., 2000 SPARC, a matricellular protein: at the crossroads of cell-matrix. Matrix biology : journal of the International Society for Matrix Biology **19**: 569–580.

Bronner-Fraser M., Fraser S. E., 1988 Cell lineage analysis reveals multipotency of some avian neural crest cells. Nature **335**: 161–164.

Brooks M. B., Gu W., Barnas J. L., Ray J., Ray K., 2004 A Line 1 insertion in the Factor IX gene segregates with mild hemophilia B in dogs. Mamm Genome **14**: 788–795.

Brouha B., Schustak J., Badge R. M., Lutz-Prigge S., Farley A. H., Moran J. V., Kazazian H. H., 2003 Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci USA **100**: 5280–5285.

Brouillard P., Dupont L., Helaers R., Coulie R., Tiller G. E., Peeden J., Colige A., Vikkula M., 2017 Loss of ADAMTS3 activity causes Hennekam lymphangiectasia-lymphedema syndrome 3. Hum Mol Genet **26**: 4095–4104.

Brown E. A., Dickinson P. J., Mansour T., Sturges B. K., Aguilar M., Young A. E., Korff C., Lind J., Ettinger C. L., Varon S., Pollard R., Brown C. T., Raudsepp T., Bannasch D. L., 2017 FGF4 retrogene on CFA12 is responsible for chondrodystrophy and intervertebral disc disease in dogs. Proc Natl Acad Sci USA **114**: 11476–11481.

Burgio G., 2018 Redefining mouse transgenesis with CRISPR/Cas9 genome editing technology. Genome Biol **19**: 27.

Cade B. E., Chen H., Stilp A. M., Gleason K. J., Sofer T., Ancoli-Israel S., Arens R., Bell G. I., Below J. E., Bjonnes A. C., Chun S., Conomos M. P., Evans D. S., Johnson W. C., Frazier-Wood A. C., Lane J. M., Larkin E. K., Loredo J. S., Post W. S., Ramos A.

R., Rice K., Rotter J. I., Shah N. A., Stone K. L., Taylor K. D., Thornton T. A., Tranah G. J., Wang C., Zee P. C., Hanis C. L., Sunyaev S. R., Patel S. R., Laurie C. C., Zhu X., Saxena R., Lin X., Redline S., 2016 Genetic Associations with Obstructive Sleep Apnea Traits in Hispanic/Latino Americans. Am J Respir Crit Care Med **194**: 886–897.

Callahan K. E., Hickman A. B., Jones C. E., Ghirlando R., Furano A. V., 2011 Polymerization and nucleic acid-binding properties of human L1 ORF1 protein. Nucleic Acids Res **40**: 813–827.

Calloni G. W., Le Douarin N. M., Dupin E., 2009 High frequency of cephalic neural crest cells shows coexistence of neurogenic, melanogenic, and osteogenic differentiation capacities. Proc Natl Acad Sci USA **106**: 8947–8952.

Campbell G. M., Sophocleous A., 2014 Quantitative analysis of bone and soft tissue by micro-computed tomography: applications to ex vivo and in vivo studies. Bonekey Rep **3**: 564.

Cantatore M., Gobbetti M., Romussi S., Brambilla G., Giudice C., Grieco V., Stefanello D., 2012 Medium term endoscopic assessment of the surgical outcome following laryngeal saccule resection in brachycephalic dogs. Vet Rec **170**: 518.

Capistrano A., Cordeiro A., Filho L. C., Almeida V. C., de Castro E Silva P. I., Martinez S., de Almeida-Pedrin R. R., 2015 Facial morphology and obstructive sleep apnea. Dental Press J Orthod **20**: 60–67.

Carlson B. M. (Ed.), 2009 Head and Neck. In: *Human Embryology and Developmental Biology*, pp. 325–360.

Carson E. A., 2006 Maximum likelihood estimation of human craniometric heritabilities. Am J Phys Anthropol **131**: 169–180.

Casinos A., Bou J., Castiella M. J., Viladiu C., 1986 On the allometry of long bones in dogs (Canis familiaris). J Morphol **190**: 73–79.

Chai Y., Jiang X., Ito Y., Bringas P., Han J., Rowitch D. H., Soriano P., McMahon A. P., Sucov H. M., 2000 Fate of the mammalian cranial neural crest during tooth and mandibular morphogenesis. Development **127**: 1671–1679.

Chambers M. G., Kuffner T., Cowan S. K., Cheah K. S. E., Mason R. M., 2002 Expression of collagen and aggrecan genes in normal and osteoarthritic murine knee joints. Osteoarthr Cartil **10**: 51–61.

Chang C. C., Chow C. C., Tellier L. C., Vattikuti S., Purcell S. M., Lee J. J., 2015 Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience **4**: 7.

Chase K., Carrier D. R., Adler F. R., Jarvik T., Ostrander E. A., Lorentzen T. D., Lark K. G., 2002 Genetic basis for systems of skeletal quantitative traits: principal component analysis of the canid skeleton. Proc Natl Acad Sci USA **99**: 9930–9935.

Chen H., Ghori-Javed F. Y., Rashid H., Adhami M. D., Serra R., Gutierrez S. E., Javed A., 2014 Runx2 regulates endochondral ossification through control of chondrocyte proliferation and differentiation. J Bone Miner Res **29**: 2653–2665.

Choi Y., Ishiguro N., Shinagawa M., Kim C. J., Okamoto Y., Minami S., Ogihara K., 1999 Molecular structure of canine LINE-1 elements in canine transmissible venereal tumor. Animal Genetics **30**: 51–53.

Chokdeemboon C., Mahatumarat C., Rojvachiranonda N., Tongkobpetch S., Suphapeetiporn K., Shotelersuk V., 2013 FGFR1 and FGFR2 mutations in Pfeiffer syndrome. J Craniofac Surg **24**: 150–152.

Cingolani P., Patel V. M., Coon M., Nguyen T., Land S. J., Ruden D. M., Lu X., 2012 Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet **3**: 35.

Cirulli V., Yebra M., 2007 Netrins: beyond the brain. Nat Rev Mol Cell Biol **8**: 296–306.

Cole J. B., Manyama M., Kimwaga E., Mathayo J., Larson J. R., Liberton D. K., Lukowiak K., Ferrara T. M., Riccardi S. L., Li M., Mio W., Prochazkova M., Williams T., Li H., Jones K. L., Klein O. D., Santorico S. A., Spritz R. A., 2016 Genomewide Association Study of African Children Identifies Association of SCHIP1 and PDE8A with Facial Size and Shape. PLoS Genetics **12**: e1006174.

Colige A., Ruggiero F., Vandenberghe I., Dubail J., Kesteloot F., Van Beeumen J., Beschin A., Brys L., Nusgens B., 2005 Domains and maturation processes that regulate the activity of ADAMTS-2, a metalloproteinase cleaving the aminopropeptide of fibrillar procollagens types I-III and V. Journal of Biological Chemistry **280**: 34397–34408.

Colliton R. P., Bason L., Lu F. M., Piccoli D. A., Krantz I. D., Spinner N. B., 2001 Mutation analysis of Jagged1 (JAG1) in Alagille syndrome patients. Hum Mutat **17**: 151–152.

Collmann H., Sörensen N., Krauss J., 2005 Hydrocephalus in craniosynostosis: a review. Childs Nerv Syst **21**: 902–912.

Collyer M.L., Sekora D.J., Adams D.C., 2015 A method for analysis of phenotypic change for phenotypes described by high-dimensionality data. Heredity. **115**: 357-365

Cooper K. L., Oh S., Sung Y., Dasari R. R., Kirschner M. W., Tabin C. J., 2013 Multiple phases of chondrocyte enlargement underlie differences in skeletal proportions. Nature **495**: 375–378.

Coppinger R., Coppinger L., 2001 *Dogs: A Startling New Understanding of Canine Origin, Behavior and Evolution*. Scribner, New York.

Couly G. F., Coltey P. M., Le Douarin N. M., 1993 The triple origin of skull in higher vertebrates: a study in quail-chick chimeras. Development **117**: 409–429.

Cox T. C., Horst J. A., Hing A. V., Cunningham M. L., 2014 Molecular Genetics and Biology of Craniosynostoses. In:Vishwakarma A, Sharpe P, Shi S, Ramalingam M (Eds.), *Stem Cell Biology and Tissue Engineering in Dental Sciences*, pp. 499–520.

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**: 265–272.

Crouch DJM., Winney B., Christmas WJ., Hutnik K., Day T., Meena D., Boumertit A., Hysi P., Nessa A., Spector TD., Kittler J., Bodmer WF., 2018 Genetics of the

human face: Identification of large-effect single gene variants. Proc Natl Acad Sci USA **115**: E676-E685.

Cruz F., Vilà C., Webster M. T., 2008 The legacy of domestication: accumulation of deleterious mutations in the dog genome. Mol Biol Evol **25**: 2331–2336.

Cui D., Li F., Li Q., Li J., Zhao Y., Hu X., Zhang R., Li N., 2015 Generation of a miniature pig disease model for human Laron syndrome. Sci Rep **5**: 15603.

Dai N., Rapley J., Angel M., Yanik M. F., Blower M. D., Avruch J., 2011 mTOR phosphorylates IMP2 to promote IGF2 mRNA translation by internal ribosomal entry. Genes Dev **25**: 1159–1172.

Dai N., Zhao L., Wrighting D., Krämer D., Majithia A., Wang Y., Cracan V., Borges-Rivera D., Mootha V. K., Nahrendorf M., Thorburn D. R., Minichiello L., Altshuler D., Avruch J., 2015 IGF2BP2/IMP2-Deficient mice resist obesity through enhanced translation of Ucp1 mRNA and Other mRNAs encoding mitochondrial proteins. Cell Metab **21**: 609–621.

Dantas F. P. M., Medeiros G. X., Figueiredo A. P. M., Thompson K., Riet-Correa F., 2013 Skeletal dysplasia with craniofacial deformity and disproportionate dwarfism in hair sheep of northeastern Brazil. J Comp Pathol **150**: 245–252.

de Crombrugghe B., Lefebvre V., Behringer R. R., Bi W., Murakami S., Huang W., 2000 Transcriptional mechanisms of chondrocyte differentiation. Matrix biology : journal of the International Society for Matrix Biology **19**: 389–394.

De Lorenzi D., Bertoncello D., Drigo M., 2009 Bronchial abnormalities found in a consecutive series of 40 brachycephalic dogs. J Am Vet Med Assoc **235**: 835–840.

de Souza M. C. C. M. I., Ryan R., Haar ter G., Packer R. M. A., Volk H. A., De Decker S., 2018 Evaluation of the influence of kyphosis and scoliosis on intervertebral disc extrusion in French bulldogs. BMC Vet Res **14**: 5.

DeLise A. M., Fischer L., Tuan R. S., 2000 Cellular interactions and signaling in cartilage development. Osteoarthr Cartil **8**: 309–334.

DePristo M. A., Banks E., Poplin R., Garimella K. V., Maguire J. R., Hartl C., Philippakis A. A., del Angel G., Rivas M. A., Hanna M., McKenna A., Fennell T. J., Kernytsky A. M., Sivachenko A. Y., Cibulskis K., Gabriel S. B., Altshuler D., Daly M. J., 2011 A

framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet **43**: 491–498.

Delaneau O., Howie B., Cox A. J., Marchini J., 2013 Haplotype estimation using sequencing reads. Am J Hum Genet **93**: 687–696.

Delaneau O., Zagury J.-F., Marchini J., 2012 Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods **10**: 5.

Dempsey J. A., Skatrud J. B., Jacques A. J., Ewanowski S. J., Woodson B. T., Hanson P. R., Goodman B., 2002 Anatomic determinants of sleep-disordered breathing across the spectrum of clinical and nonclinical male subjects. Chest **122**: 840–851.

Deng J., Murphy B. A., Dietrich M. S., Wells N., Wallston K. A., Sinard R. J., Cmelak A. J., Gilbert J., Ridner S. H., 2012 Impact of secondary lymphedema after head and neck cancer treatment on symptoms, functional status, and quality of life. Head Neck **35**: 1026–1035.

Di Francesco R., Monteiro R., de Melo Paulo M. L., Buranello F., Imamura R., 2012 Craniofacial morphology and sleep apnea in children with obstructed upper airways: differences between genders. Sleep Med **13**: 616–620.

Dierks C., Mömke S., Philipp U., Distl O., 2013 Allelic heterogeneity of FGF5 mutations causes the long-hair phenotype in dogs. Animal Genetics **44**: 425–431.

Dietschi E., Ruchti M., Gaillard C., Stich H., Schawalder P., 2010 Oberes Luftweg-Syndrom beim Norwich Terrier. Zeitschrift der Schweizerischen Kynologischen Gesellschaft SKG vom 19 **3**.

Dobin A., Davis C. A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T. R., 2012 STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**: 15–21.

Dottori M., Gross M. K., Labosky P., Goulding M., 2001 The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. Development **128**: 4127–4138.

Downs L. M., Mellersh C. S., 2014 An Intronic SINE insertion in FAM161A that causes exon-skipping is associated with progressive retinal atrophy in Tibetan Spaniels and Tibetan Terriers. PLoS ONE **9**: e93990.

Drake A. G., 2011 Dispelling dog dogma: an investigation of heterochrony in dogs using 3D geometric morphometric analysis of skull shape. Evol Dev **13**: 204–213.

Drake A. G., Klingenberg C. P., 2007 The pace of morphological change: historical transformation of skull shape in St Bernard dogs. Proc Biol Sci **275**: 71–76.

Drake A. G., Klingenberg C. P., 2010 Large-scale diversification of skull shape in domestic dogs: disparity and modularity. Am Nat **175**: 289–301.

Drake A. G., Coquerelle M., Kosintsev P. A., Bachura O. P., Sablin M., Gusev A. V., Fleming L. S., Losey R. J., 2017 Three-Dimensional Geometric Morphometric Analysis of Fossil Canid Mandibles and Skulls. Sci Rep **7**: 9508.

Drake A. G., Coquerelle M., Colombeau G., 2015 3D morphometric analysis of fossil canid skulls contradicts the suggested domestication of dogs during the late Paleolithic. Sci Rep **5**: 8299.

Dreger D. L., Schmutz S. M., 2011 A SINE insertion causes the black-and-tan and saddle tan phenotypes in domestic dogs. J Hered **102 Suppl 1**: S11–8.

Driessen C., Joosten K. F. M., Bannink N., Bredero-Boelhouwer H. H., Hoeve H. L. J., Wolvius E. B., Rizopoulos D., Mathijssen I. M. J., 2013 How does obstructive sleep apnoea evolve in syndromic craniosynostosis? A prospective cohort study. Arch Dis Child **98**: 538–543.

Drögemüller C., Giese A., Martins-Wess F., Wiedemann S., Andersson L., Brenig B., Fries R., Leeb T., 2006 The mutation causing the black-and-tan pigmentation phenotype of Mangalitza pigs maps to the porcine ASIP locus but does not affect its coding sequence. Mamm Genome **17**: 58–66.

Ellis J. L., Thomason J., Kebreab E., Zubair K., France J., 2009 Cranial dimensions and forces of biting in the domestic dog. J Anat **214**: 362–373.

Epstein D. J., Vekemans M., Gros P., 1991 Splotch (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. Cell **67**: 767–774.

Erickson C. A., Weston J. A., 1983 An SEM analysis of neural crest migration in the mouse. J Embryol Exp Morphol **74**: 97–118.

Eriksson K. S., Mignot E., 2009 T-box 3 is expressed in the adult mouse hypothalamus and medulla. Brain Res **1302**: 233–239.

Erlebacher A., Filvaroff E. H., Gitelman S. E., Derynck R., 1995 Toward a molecular understanding of skeletal development. Cell **80**: 371–378.

Estécio M. R. H., Gallegos J., Dekmezian M., Lu Y., Liang S., Issa J.-P. J., 2012 SINE retrotransposons cause epigenetic reprogramming of adjacent gene promoters. Mol Cancer Res **10**: 1332–1342.

Evans H. E., 1958 Prental Ossification in the Dog. Anat Rec (Hoboken): 130–406.

Evans H. E., 1993 The Skeleton. In:Evans HE (Ed.), *Miller's Anatomy of the Dog*, W.B. Saunders Company, Philadelphia, pp. 122–218.

Evans H. E., de Lahunta A., 2013 *Miller's Anatomy of the Dog*. Elsevier Saunders.

Everson R., Pettitt L., Forman O. P., Dower-Tylee O., McLaughlin B., Ahonen S., Kaukonen M., Komáromy A. M., Lohi H., Mellersh C. S., Sansom J., Ricketts S. L., 2017 An intronic LINE-1 insertion in MERTK is strongly associated with retinopathy in Swedish Vallhund dogs. PLoS ONE **12**: e0183021.

Fallahshahroudi A., Løtvedt P., Bélteky J., Altimiras J., Jensen P., 2018 Changes in pituitary gene expression may underlie multiple domesticated traits in chickens. Heredity (Edinb).

Fasanella F. J., Shivley J. M., Wardlaw J. L., Givaruangsawat S., 2010 Brachycephalic airway obstructive syndrome in dogs: 90 cases (1991-2008). J Am Vet Med Assoc **237**: 1048–1051.

Feng Q., Moran J. V., Kazazian H. H., Boeke J. D., 1996 Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell **87**: 905–916.

Feng W., Leach S. M., Tipney H., Phang T., Geraci M., Spritz R. A., Hunter L. E., Williams T., 2009 Spatial and temporal analysis of gene expression during growth and fusion of the mouse facial prominences. PLoS ONE **4**: e8066.

Ferguson K. A., Ono T., Lowe A. A., Ryan C. F., Fleetham J. A., 1995 The relationship between obesity and craniofacial structure in obstructive sleep apnea. Chest **108**: 375–381.

Fernandes R. J., Hirohata S., Engle J. M., Colige A., Cohn D. H., Eyre D. R., Apte S. S., 2001 Procollagen II amino propeptide processing by ADAMTS-3. Insights on dermatosparaxis. Journal of Biological Chemistry **276**: 31502–31509.

Finnerty J. R., Mazza M. E., Jezewski P. A., 2009 Domain duplication, divergence, and loss events in vertebrate Msx paralogs reveal phylogenomically informed disease markers. BMC Evol Biol **9**: 18.

Florisson J. M. G., Verkerk A. J. M. H., Huigh D., Hoogeboom A. J. M., Swagemakers S., Kremer A., Heijsman D., Lequin M. H., Mathijssen I. M. J., van der Spek P. J., 2013 Boston type craniosynostosis: report of a second mutation in MSX2. Am J Med Genet A **161A**: 2626–2633.

Fondon J. W., Garner H. R., 2004 Molecular origins of rapid and continuous morphological evolution. Proc Natl Acad Sci USA **101**: 18058–18063.

Frantz L. A. F., Mullin V. E., Pionnier-Capitan M., Lebrasseur O., Ollivier M., Perri A., Linderholm A., Mattiangeli V., Teasdale M. D., Dimopoulos E. A., Tresset A., Duffraisse M., McCormick F., Bartosiewicz L., Gál E., Nyerges É. A., Sablin M. V., Bréhard S., Mashkour M., Bălăşescu A., Gillet B., Hughes S., Chassaing O., Hitte C., Vigne J.-D., Dobney K., Hänni C., Bradley D. G., Larson G., 2016 Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science **352**: 1228–1231.

Franz-Odendaal T. A., Hall B. K., Witten P. E., 2005 Buried alive: how osteoblasts become osteocytes. Dev Dyn **235**: 176–190.

Freedman A. H., Gronau I., Schweizer R. M., Vecchyo D. O.-D., Han E., Silva P. M., Galaverni M., Fan Z., Marx P., Lorente-Galdos B., Beale H., Ramirez O., Hormozdiari F., Alkan C., Vilà C., Squire K., Geffen E., Kusak J., Boyko A. R., Parker H. G., Lee C., Tadigotla V., Wilton A., Siepel A., Bustamante C. D., Harkins T. T., Nelson S. F., Ostrander E. A., Marques-Bonet T., Wayne R. K., Novembre J., 2014

Genome sequencing highlights the dynamic early history of dogs. PLoS Genetics **10**: e1004016.

Frischknecht M., Niehof-Oellers H., Jagannathan V., Owczarek-Lipska M., Drögemüller C., Dietschi E., Dolf G., Tellhelm B., Lang J., Tiira K., Lohi H., Leeb T., 2013 A COL11A2 mutation in Labrador retrievers with mild disproportionate dwarfism. PLoS ONE **8**: e60149.

Frynta D., Baudyšová J., Hradcová P., Faltusová K., Kratochvíl L., 2012 Allometry of sexual size dimorphism in domestic dog. PLoS ONE **7**: e46125.

Fu Y., Li C., Tang Q., Tian S., Jin L., Chen J., Li M., Li C., 2016 Genomic analysis reveals selection in Chinese native black pig. Sci Rep **6**: 36354.

Fuchsberger C., Abecasis G. R., Hinds D. A., 2014 minimac2: faster genotype imputation. Bioinformatics **31**: 782–784.

Fusco I., Babu D., Mellone S., Barizzone N., Prodam F., Fanelli A., Muniswamy R., Petri A., Bellone S., Bona G., Giordano M., 2015 Variations in the high-mobility group-A2 gene (HMGA2) are associated with idiopathic short stature. Pediatr Res **79**: 258–261.

Gagliardi A. D., Kuo E. Y. W., Raulic S., Wagner G. F., DiMattia G. E., 2004 Human stanniocalcin-2 exhibits potent growth-suppressive properties in transgenic mice independently of growth hormone and IGFs. Am J Physiol Endocrinol Metab **288**: E92–105.

Geiger M., Haussman S., 2016 Cranial Suture Closure in Domestic Dog Breeds and Its Relationships to Skull Morphology. Anat Rec (Hoboken) **299**: 412–420.

Geiger M., Evin A., Sánchez-Villagra M. R., Gascho D., Mainini C., Zollikofer C. P. E., 2017 Neomorphosis and heterochrony of skull shape in dog domestication. Sci Rep **7**: 13443.

Germonpré M., Lázničková-Galetová M., Sablin M. V., 2012 Palaeolithic dog skulls at the Gravettian Předmostí site, the Czech Republic. Journal of Archaeological Science **39**: 184.

Germonpré M., Sablin M. V., Stevens R. E., Hofreiter M., Stiller M., Després V. R., 2009 Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and

Russia: osteometry, ancient DNA and stable isotopes. Journal of Archaeological Science **36**: 473.

Ghouzzi el V., Le Merrer M., Perrin-Schmitt F., Lajeunie E., Benit P., Renier D., Bourgeois P., Bolcato-Bellemin A. L., Munnich A., Bonaventure J., 1997 Mutations of the TWIST gene in the Saethre-Chotzen syndrome. Nat Genet **15**: 42–46.

Gibson G., 1998 Active role of chondrocyte apoptosis in endochondral ossification. Microsc Res Tech **43**: 191–204.

Gilbert N., Lutz S., Morrish T. A., Moran J. V., 2005 Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol Cell Biol **25**: 7780–7795.

Ginn J. A., Kumar M. S. A., McKiernan B. C., Powers B. E., 2008 Nasopharyngeal turbinates in brachycephalic dogs and cats. J Am Anim Hosp Assoc **44**: 243–249.

Girardot M., Martin J., Guibert S., Leveziel H., Julien R., Oulmouden A., 2005 Widespread expression of the bovine Agouti gene results from at least three alternative promoters. Pigment Cell Res **18**: 34–41.

Goldstein O., Guyon R., Kukekova A., Kuznetsova T. N., Pearce-Kelling S. E., Johnson J., Aguirre G. D., Acland G. M., 2010 COL9A2 and COL9A3 mutations in canine autosomal recessive oculoskeletal dysplasia. Mamm Genome **21**: 398–408.

Gould S. J., 1973 Positive Allometry of Antlers in the “Irish Elk,” Megaloceros giganteus. Nature **244**: 375–376.

Goulding M. D., Chalepakis G., Deutsch U., Erselius J. R., Gruss P., 1991 Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. EMBO J **10**: 1135–1147.

Grilo A., Ruiz-Granados E. S., Moreno-Rey C., Rivera J. M., Ruiz A., Real L. M., Sáez M. E., 2013 Genetic analysis of candidate SNPs for metabolic syndrome in obstructive sleep apnea (OSA). Gene **521**: 150–154.

Guenou H., Kaabeche K., Le Mée S., Marie P. J., 2005 A role for fibroblast growth factor receptor-2 in the altered osteoblast phenotype induced by Twist

haploinsufficiency in the Saethre-Chotzen syndrome. Hum Mol Genet **14**: 1429–1439.

Gunz P., Mitteroecker P., 2013 Semilandmarks: a method for quantifying curves and surfaces. Ital J Zool. **24**:103-109

Guo J., Mei X., Tang K., 2013 Automatic landmark annotation and dense correspondence registration for 3D human facial images. BMC Bioinformatics **14**: 232.

Göhring W., Sasaki T., Heldin C. H., Timpl R., 1998 Mapping of the binding of platelet-derived growth factor to distinct domains of the basement membrane proteins BM-40 and perlecan and distinction from the BM-40 collagen-binding epitope. Eur J Biochem **255**: 60–66.

Hagiwara K., Obayashi T., Sakayori N., Yamanishi E., Hayashi R., Osumi N., Nakazawa T., Nishida K., 2014 Molecular and cellular features of murine craniofacial and trunk neural crest cells as stem cell-like cells. PLoS ONE **9**: e84072.

Hall B. K., 2001 The neural crest as a fourth germ layer and vertebrates as quadroblastic not triploblastic. Evol Dev **2**: 3–5.

Hammond P., Hutton T. J., Allanson J. E., Campbell L. E., Hennekam R. C. M., Holden S., Patton M. A., Shaw A., Temple I. K., Trotter M., Murphy K. C., Winter R. M., 2004 3D analysis of facial morphology. Am J Med Genet **126A**: 339–348.

Han J. S., Szak S. T., Boeke J. D., 2004 Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature **429**: 268–274.

Harvey C. E., 1989 Inherited and Congenital Airway Conditions. J Small Anim Pract **30**: 184–187.

Harvey R. G., Haar ter G., 2016 Brachycephalic Obstructive Airway Syndrome. In:Haar ter G, Harvey RG (Eds.), *Ear, Nose and Throat Diseases of the Dog and Cat*, CRC Press, Devon, UK, pp. 288–292.

Hayashi S., Kim J. H., Hwang S. E., Shibata S., Fujimiya M., Murakami G., Cho B. H., 2014 Interface between intramembranous and endochondral ossification in human foetuses. Folia Morphol (Warsz) **73**: 199–205.

Hayashi Y., Yamamoto N., Nakagawa T., Ito J., 2013 Insulin-like growth factor 1 induces the transcription of Gap43 and Ntn1 during hair cell protection in the neonatal murine cochlea. Neurosci Lett **560**: 7–11.

Hayward J. J., Castelhano M. G., Oliveira K. C., Corey E., Balkman C., Baxter T. L., Casal M. L., Sharon A Center, Fang M., Garrison S. J., Kalla S. E., Korniliev P., Kotlikoff M. I., Moise N. S., Shannon L. M., Simpson K. W., Sutter N. B., Todhunter R. J., Boyko A. R., 2016 Complex disease and phenotype mapping in the domestic dog. Nat Commun **7**: 10460.

Healy C., Uwanogho D., Sharpe P. T., 1999 Regulation and role of Sox9 in cartilage formation. Dev Dyn **215**: 69–78.

Hebrok M., Reichardt L. F., 2004 Brain meets pancreas: netrin, an axon guidance molecule, controls epithelial cell migration. Trends Cell Biol **14**: 153–155.

Heidenreich D., Gradner G., Kneissl S., Dupré G., 2016 Nasopharyngeal Dimensions From Computed Tomography of Pugs and French Bulldogs With Brachycephalic Airway Syndrome. Vet Surg **45**: 83–90.

Hekman J. P., Johnson J. L., Edwards W., Vladimirova A. V., Gulevich R. G., Ford A. L., Kharlamova A. V., Herbeck Y., Acland G. M., Raetzman L. T., Trut L. N., Kukekova A. V., 2018 Anterior Pituitary Transcriptome Suggests Differences in ACTH Release in Tame and Aggressive Foxes. G3 (Bethesda) **8**: 859–873.

Helms J. A., Cordero D., Tapadia M. D., 2005 New insights into craniofacial morphogenesis. Development **132**: 851–861.

Helmsmüller D., Wefstaedt P., Nolte I., Schilling N., 2013 Ontogenetic allometry of the Beagle. BMC Vet Res **9**: 203.

Henderson D. J., Copp A. J., 1998 Role of the extracellular matrix in neural crest cell migration. J Anat **191 (Pt 4)**: 507–515.

Hillbertz N. H. C. S., Isaksson M., Karlsson E. K., Hellmén E., Pielberg G. R., Savolainen P., Wade C. M., Euler von H., Gustafson U., Hedhammar A., Nilsson M., Lindblad-Toh K., Andersson L., Andersson G., 2007 Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. Nat Genet **39**: 1318–1320.

Hodgman S.F.J., 1963 Abnormalities and defects in pedigree dogs 1. An investigation into the existence of abnormalities in pedigree dogs in the British Isles. J Small Anim Pract: 447–456.

Hohjoh H., Singer M. F., 1997 Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. EMBO J **16**: 6034–6043.

Hoopes B. C., Rimbault M., Liebers D., Ostrander E. A., Sutter N. B., 2012 The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. Mamm Genome **23**: 780–790.

Housley D. J. E., Venta P. J., 2006 The long and the short of it: evidence that FGF5 is a major determinant of canine "hair-"itability. Animal Genetics **37**: 309–315.

Howard T. D., Paznekas W. A., Green E. D., Chiang L. C., Ma N., de Luna R. I. O., Delgado C. G., Gonzalez-Ramos M., Kline A. D., Jabs E. W., 1997 Mutations in TWIST, a basic helix-loop-helix transcription factor, in Saethre-Chotzen syndrome. Nat Genet **15**: 36–41.

Huber W., 1974 Biometric Analysis of Brachycephaly in Domestic Dogs. Animal Biology: 3–4.

Hunter W. S., Balbach D. R., Lamphiear D. E., 1970 The heritability of attained growth in the human face. Am J Orthod **58**: 128–134.

Huszar D., Lynch C. A., Fairchild-Huntress V., Dunmore J. H., Fang Q., Berkemeier L. R., Gu W., Kesterson R. A., Boston B. A., Cone R. D., Smith F. J., Campfield L. A., Burn P., Lee F., 1997 Targeted disruption of the melanocortin-4 receptor results in obesity in mice. Cell **88**: 131–141.

Jabalee J., Hillier S., Franz-Odendaal T. A., 2013 An investigation of cellular dynamics during the development of intramembranous bones: the scleral ossicles. J Anat **223**: 311–320.

Jabs E. W., Li X., Scott A. F., Meyers G., Chen W., Eccles M., Mao J. I., Charnas L. R., Jackson C. E., Jaye M., 1994 Jackson-Weiss and Crouzon syndromes are allelic with mutations in fibroblast growth factor receptor 2. Nat Genet **8**: 275–279.

Jabs E. W., Müller U., Li X., Ma L., Luo W., Haworth I. S., Klisak I., Sparkes R., Warman M. L., Mulliken J. B., 1993 A mutation in the homeodomain of the

human MSX2 gene in a family affected with autosomal dominant craniosynostosis. Cell **75**: 443–450.

Janners M. Y., Searls R. L., 1970 Changes in rate of cellular proliferation during the differentiation of cartilage and muscle in the mesenchyme of the embryonic chick wing. Dev Biol **23**: 136–165.

Jeltsch M., Jha S. K., Tvorogov D., Anisimov A., Leppänen V.-M., Holopainen T., Kivelä R., Ortega S., Kärpanen T., Alitalo K., 2014 CCBE1 enhances lymphangiogenesis via A disintegrin and metalloprotease with thrombospondin motifs-3-mediated vascular endothelial growth factor-C activation. Circulation **129**: 1962–1971.

Jha S. K., Rauniyar K., Leppänen V.-M., Brouillard P., Vikkula M., Alitalo K., Jeltsch M., 2017 Efficient activation of the lymphangiogenic growth factor VEGF-C requires the C-terminal domain of VEGF-C and the N-terminal domain of CCBE1. Sci Rep **7**: 4916.

Jiang X., Iseki S., Maxson R. E., Sucov H. M., Morriss-Kay G. M., 2002 Tissue origins and interactions in the mammalian skull vault. Dev Biol **241**: 106–116.

Johannsdottir B., Thorarinsson F., Thordarson A., Magnusson T. E., 2005 Heritability of craniofacial characteristics between parents and offspring estimated from lateral cephalograms. Am J Orthod Dentofacial Orthop **127**: 200.

Johnson D., Wilkie A. O. M., 2011 Craniosynostosis. Eur J Hum Genet **19**: 369–376.

Johnson L. R., Mayhew P. D., Steffey M. A., Hunt G. B., Carr A. H., McKiernan B. C., 2013 Upper airway obstruction in Norwich Terriers: 16 cases. J Vet Intern Med **27**: 1409–1415.

Johnston M. C., 1966 A radioautographic study of the migration and fate of cranial neural crest cells in the chick embryo. Anat Rec (Hoboken) **156**: 143–155.

Jones P., Chase K., Martin A., Davern P., Ostrander E. A., Lark K. G., 2008 Single-nucleotide-polymorphism-based association mapping of dog stereotypes. Genetics **179**: 1033–1044.

Jurka J., 1997 Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proceedings of the National Academy of Sciences of the United States of America **94**: 1872–1877.

Justice C. M., Kim J., Kim S.-D., Kim K., Yagnik G., Cuellar A., Carrington B., Lu C.-L., Sood R., Boyadjiev S. A., Wilson A. F., 2017 A variant associated with sagittal nonsyndromic craniosynostosis alters the regulatory function of a non-coding element. Am J Med Genet A **173**: 2893–2897.

Justice C. M., Yagnik G., Kim Y., Peter I., Jabs E. W., Erazo M., Ye X., Ainehsazan E., Shi L., Cunningham M. L., Kimonis V., Roscioli T., Wall S. A., Wilkie A. O. M., Stoler J., Richtsmeier J. T., Heuzé Y., Sanchez-Lara P. A., Buckley M. F., Druschel C. M., Mills J. L., Caggana M., Romitti P. A., Kay D. M., Senders C., Taub P. J., Klein O. D., Boggan J., Zwienenberg-Lee M., Naydenov C., Kim J., Wilson A. F., Boyadjiev S. A., 2012 A genome-wide association study identifies susceptibility loci for nonsyndromic sagittal craniosynostosis near BMP2 and within BBS9. Nat Genet **44**: 1360–1364.

Kalff-Suske M., Wild A., Topp J., Wessling M., Jacobsen E. M., Bornholdt D., Engel H., Heuer H., Aalfs C. M., Ausems M. G., Barone R., Herzog A., Heutink P., Homfray T., Gillessen-Kaesbach G., König R., Kunze J., Meinecke P., Müller D., Rizzo R., Strenge S., Superti-Furga A., Grzeschik K. H., 1999 Point mutations throughout the GLI3 gene cause Greig cephalopolysyndactyly syndrome. Hum Mol Genet **8**: 1769–1777.

Kaminska D., Käkelä P., Nikkola E., Venesmaa S., Ilves I., Herzig K.-H., Kolehmainen M., Karhunen L., Kuusisto J., Gylling H., Pajukanta P., Laakso M., Pihlajamäki J., 2016 Regulation of alternative splicing in human obesity loci. Obesity (Silver Spring) **24**: 2033–2037.

Kang J. S., Alliston T., Delston R., Derynck R., 2005 Repression of Runx2 function by TGF-beta through recruitment of class II histone deacetylases by Smad3. EMBO J **24**: 2543–2555.

Karyadi D. M., Karlins E., Decker B., vonHoldt B. M., Carpintero-Ramirez G., Parker H. G., Wayne R. K., Ostrander E. A., 2013 A copy number variant at the KITLG

locus likely confers risk for canine squamous cell carcinoma of the digit. PLoS Genetics **9**: e1003409.

Katschenko N., 1888 Zur Entwicklungsgeschichte der Selachier-embryos. Anatomische Anziger: 445–467.

Kazazian H. H., Wong C., Youssoufian H., Scott A. F., Phillips D. G., Antonarakis S. E., 1988 Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature **332**: 164–166.

Khanna C., Lindblad-Toh K., Vail D., London C., Bergman P., Barber L., Breen M., Kitchell B., McNeil E., Modiano J. F., Niemi S., Comstock K. E., Ostrander E., Westmoreland S., Withrow S., 2006 The dog as a cancer model. Nat Biotechnol **24**: 1065–1066.

Kikuchi M., Higurashi N., Miyazaki S., Itasaka Y., 2001 Facial patterns of obstructive sleep apnea patients using Ricketts' method. Psychiatry Clin Neurosci **54**: 336–337.

Kimonis V., Gold J.-A., Hoffman T. L., Panchal J., Boyadjiev S. A., 2007 Genetics of craniosynostosis. Semin Pediatr Neurol **14**: 150–161.

King L., Harris E. F., Tolley E. A., 1993 Heritability of cephalometric and occlusal variables as assessed from siblings with overt malocclusions. Am J Orthod Dentofacial Orthop **104**: 121–131.

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**: 1898–1903.

Klingenberg C. P., 2010 MorphoJ: an integrated software package for geometric morphometrics. Mol Ecol Resour **11**: 353–357.

Klingenberg C. P., 2016 Size, shape, and form: concepts of allometry in geometric morphometrics. Dev Genes Evol **226**: 113–137.

Knowler S. P., Cross C., Griffiths S., McFadyen A. K., Jovanovik J., Tauro A., Kibar Z., Driver C. J., La Ragione R. M., Rusbridge C., 2017 Use of Morphometric Mapping to Characterise Symptomatic Chiari-Like Malformation, Secondary Syringomyelia

and Associated Brachycephaly in the Cavalier King Charles Spaniel. PLoS ONE **12**: e0170315.

Koch D. A., Arnold S., Hubler M., Montavon P. M., 2003 Brachycephalic Syndrome in Dogs. Compendium on Continuing Education for the Practising Veterinarian **25**: 48–55.

Koch D., Wiestner T., Balli A., Montavon P., Michel E., Scharf G., Arnold S., 2012 Proposal for a new radiological index to determine skull conformation in the dog. Schweiz Arch Tierheilkd **154**: 217–220.

Koch I. J., Clark M. M., Thompson M. J., Deere-Machemer K. A., Wang J., Duarte L., Gnanadesikan G. E., McCoy E. L., Rubbi L., Stahler D. R., Pellegrini M., Ostrander E. A., Wayne R. K., Sinsheimer J. S., vonHoldt B. M., 2016 The concerted impact of domestication and transposon insertions on methylation patterns between dogs and grey wolves. Mol Ecol **25**: 1838–1855.

Koh I., Lee M. S., Lee M. S., Lee N. J., Park K. W., Kim K. H., Kim H., Rhyu I. J., 2005 Body size effect on brain volume in Korean youth. Neuroreport **16**: 2029–2032.

Kong L.-L., Yang N.-Z., Shi L.-H., Zhao G.-H., Zhou W., Ding Q., Wang M.-H., Zhang Y.-S., 2017 The optimum marker for the detection of lymphatic vessels. Mol Clin Oncol **7**: 515–520.

Kopera H. C., Larson P. A., Moldovan J. B., Richardson S. R., Liu Y., Moran J. V., 2016 LINE-1 Cultured Cell Retrotransposition Assay. Methods Mol Biol **1400**: 139–156.

Koyabu D., Maier W., Sánchez-Villagra M. R., 2012 Paleontological and developmental evidence resolve the homology and dual embryonic origin of a mammalian skull bone, the interparietal. Proc Natl Acad Sci USA **109**: 14075–14080.

Krakow D., 2015 Skeletal Dysplasias. Clinics in Perinatology **42**: 301.

Kripke D. F., Kline L. E., Nievergelt C. M., Murray S. S., Shadan F. F., Dawson A., Poceta J. S., Cronin J., Jamil S. M., Tranah G. J., Loving R. T., Grizas A. P., Hahn E. K., 2014 Genetic variants associated with sleep disorders. Sleep Med **16**: 217–224.

Kristensen E., Parsons T. E., Hallgrímsson B., Boyd S. K., 2009 A novel 3-D image-based morphological method for phenotypic analysis. IEEE Trans Biomed Eng **55**: 2826–2831.

Kupprion C., Motamed K., Sage E. H., 1998 SPARC (BM-40, osteonectin) inhibits the mitogenic effect of vascular endothelial growth factor on microvascular endothelial cells. Journal of Biological Chemistry **273**: 29635–29640.

Kyöstilä K., Lappalainen A. K., Lohi H., 2013 Canine chondrodysplasia caused by a truncating mutation in collagen-binding integrin alpha subunit 10. PLoS ONE **8**: e75621.

Körberg I. B., Sundström E., Meadows J. R. S., Pielberg G. R., Gustafson U., Karlsson E. K., Seddon J., Söderberg A., Vilà C., Zhang X., Åkesson M., Lindblad-Toh K., Andersson G., Andersson L., 2014 A simple repeat polymorphism in the MITF-M promoter is a key regulator of white spotting in dogs. PLoS ONE **9**: e104363.

Künzel W., Breit S., Oppel M., 2003 Morphometric investigations of breed-specific features in feline skulls and considerations on their functional implications. Anat Histol Embryol **32**: 218–223.

Ladlow J., Liu N.-C., Kalmar L., Sargan D., 2018 Brachycephalic obstructive airway syndrome. Vet Rec **182**: 375–378.

Lajeunie E., Le Merrer M., Bonaïti-Pellie C., Marchac D., Renier D., 1995 Genetic study of nonsyndromic coronal craniosynostosis. Am J Med Genet **55**: 500–504.

Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., Funke R., Gage D., Harris K., Heaford A., Howland J., Kann L., Lehoczky J., LeVine R., McEwan P., McKernan K., Meldrim J., Mesirov J. P., Miranda C., Morris W., Naylor J., Raymond C., Rosetti M., Santos R., Sheridan A., Sougnez C., Stange-Thomann Y., Stojanovic N., Subramanian A., Wyman D., Rogers J., Sulston J., Ainscough R., Beck S., Bentley D., Burton J., Clee C., Carter N., Coulson A., Deadman R., Deloukas P., Dunham A., Dunham I., Durbin R., French L., Grafham D., Gregory S., Hubbard T., Humphray S., Hunt A., Jones M., Lloyd C., McMurray A., Matthews L., Mercer S., Milne S., Mullikin J. C., Mungall A., Plumb R., Ross M., Shownkeen R., Sims S., Waterston R. H., Wilson R.

K., Hillier L. W., McPherson J. D., Marra M. A., Mardis E. R., Fulton L. A., Chinwalla A. T., Pepin K. H., Gish W. R., Chissoe S. L., Wendl M. C., Delehaunty K. D., Miner T. L., Delehaunty A., Kramer J. B., Cook L. L., Fulton R. S., Johnson D. L., Minx P. J., Clifton S. W., Hawkins T., Branscomb E., Predki P., Richardson P., Wenning S., Slezak T., Doggett N., Cheng J. F., Olsen A., Lucas S., Elkin C., Uberbacher E., Frazier M., Gibbs R. A., Muzny D. M., Scherer S. E., Bouck J. B., Sodergren E. J., Worley K. C., Rives C. M., Gorrell J. H., Metzker M. L., Naylor S. L., Kucherlapati R. S., Nelson D. L., Weinstock G. M., Sakaki Y., Fujiyama A., Hattori M., Yada T., Toyoda A., Itoh T., Kawagoe C., Watanabe H., Totoki Y., Taylor T., Weissenbach J., Heilig R., Saurin W., Artiguenave F., Brottier P., Bruls T., Pelletier E., Robert C., Wincker P., Smith D. R., Doucette-Stamm L., Rubenfield M., Weinstock K., Lee H. M., Dubois J., Rosenthal A., Platzer M., Nyakatura G., Taudien S., Rump A., Yang H., Yu J., Wang J., Huang G., Gu J., Hood L., Rowen L., Madan A., Qin S., Davis R. W., Federspiel N. A., Abola A. P., Proctor M. J., Myers R. M., Schmutz J., Dickson M., Grimwood J., Cox D. R., Olson M. V., Kaul R., Raymond C., Shimizu N., Kawasaki K., Minoshima S., Evans G. A., Athanasiou M., Schultz R., Roe B. A., Chen F., Pan H., Ramser J., Lehrach H., Reinhardt R., McCombie W. R., la Bastide de M., Dedhia N., Blöcker H., Hornischer K., Nordsiek G., Agarwala R., Aravind L., Bailey J. A., Bateman A., Batzoglou S., Birney E., Bork P., Brown D. G., Burge C. B., Cerutti L., Chen H. C., Church D., Clamp M., Copley R. R., Doerks T., Eddy S. R., Eichler E. E., Furey T. S., Galagan J., Gilbert J. G., Harmon C., Hayashizaki Y., Haussler D., Hermjakob H., Hokamp K., Jang W., Johnson L. S., Jones T. A., Kasif S., Kaspryzk A., Kennedy S., Kent W. J., Kitts P., Koonin E. V., Korf I., Kulp D., Lancet D., Lowe T. M., McLysaght A., Mikkelsen T., Moran J. V., Mulder N., Pollara V. J., Ponting C. P., Schuler G., Schultz J., Slater G., Smit A. F., Stupka E., Szustakowki J., Thierry-Mieg D., Thierry-Mieg J., Wagner L., Wallis J., Wheeler R., Williams A., Wolf Y. I., Wolfe K. H., Yang S. P., Yeh R. F., Collins F., Guyer M. S., Peterson J., Felsenfeld A., Wetterstrand K. A., Patrinos A., Morgan M. J., de Jong P., Catanese J. J., Osoegawa K., Shizuya H., Choi S., Chen Y. J., Szustakowki J.,

2001 Initial sequencing and analysis of the human genome. Nature **409**: 860–921.

Larkin E. K., Patel S. R., Goodloe R. J., Li Y., Zhu X., Gray-McGuire C., Adams M. D., Redline S., 2010 A candidate gene study of obstructive sleep apnea in European Americans and African Americans. Am J Respir Crit Care Med **182**: 947–953.

Larson G., Karlsson E. K., Perri A., Webster M. T., Ho S. Y. W., Peters J., Stahl P. W., Piper P. J., Lingaas F., Fredholm M., Comstock K. E., Modiano J. F., Schelling C., Agoulnik A. I., Leegwater P. A., Dobney K., Vigne J.-D., Vilà C., Andersson L., Lindblad-Toh K., 2012 Rethinking dog domestication by integrating genetics, archeology, and biogeography. Proc Natl Acad Sci USA **109**: 8878–8883.

Lattimore J.-D. L., Celermajer D. S., Wilcox I., 2003 Obstructive sleep apnea and cardiovascular disease. J Am Coll Cardiol **41**: 1429–1437.

Le Douarin N., 1973 A biological cell labeling technique and its use in experimental embryology. Dev Biol **30**: 217.

Le Douarin N. M., 2004 The avian embryo as a model to study the development of the neural crest: a long and still ongoing story. Mech Dev **121**: 1089–1102.

Le Goff C., Somerville R. P. T., Kesteloot F., Powell K., Birk D. E., Colige A. C., Apte S. S., 2006 Regulation of procollagen amino-propeptide processing during mouse embryogenesis by specialization of homologous ADAMTS proteases: insights on collagen biosynthesis and dermatosparaxis. Development **133**: 1587–1596.

Le Lièvre C. S., 1978 Participation of neural crest-derived cells in the genesis of the skull in birds. J Embryol Exp Morphol **47**: 17–37.

Lee M. K., Shaffer J. R., Leslie E. J., Orlova E., Carlson J. C., Feingold E., Marazita M. L., Weinberg S. M., 2017 Genome-wide association study of facial morphology reveals novel associations with FREM1 and PARK2. PLoS ONE **12**: e0176566.

Lee R. W. W., Petocz P., Prvan T., Chan A. S. L., Grunstein R. R., Cistulli P. A., 2009 Prediction of obstructive sleep apnea with craniofacial photographic analysis. J Appl Physiol **32**: 46–52.

Lenton K., James A. W., Manu A., Brugmann S. A., Birker D., Nelson E. R., Leucht P., Helms J. A., Longaker M. T., 2011 Indian hedgehog positively regulates calvarial

ossification and modulates bone morphogenetic protein signaling. Genesis **49**: 784–796.

Leslie E. J., Taub M. A., Liu H., Steinberg K. M., Koboldt D. C., Zhang Q., Carlson J. C., Hetmanski J. B., Wang H., Larson D. E., Fulton R. S., Kousa Y. A., Fakhouri W. D., Naji A., Ruczinski I., Begum F., Parker M. M., Busch T., Standley J., Rigdon J., Hecht J. T., Scott A. F., Wehby G. L., Christensen K., Czeizel A. E., Schutte B. C., Wilson R. K., Cornell R. A., Lidral A. C., Weinstock G. M., Beaty T. H., Marazita M. L., Murray J. C., 2015 Identification of functional variants for cleft lip with or without cleft palate in or near PAX7, FGFR2, and NOG by targeted sequencing of GWAS loci. Am J Hum Genet **96**: 397–411.

Lewis A. E., Vasudevan H. N., O'Neill A. K., Soriano P., Bush J. O., 2013 The widely used Wnt1-Cre transgene causes developmental phenotypes by ectopic activation of Wnt signaling. Dev Biol **379**: 229–234.

Li H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv **q-bio.GN**.

Liao Y., Smyth G. K., Shi W., 2013 The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res **41**: e108.

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., 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. J., 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., 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., Foley C., Franke A., Friedrich D., Gage D., Garber M., Gearin G., Giannoukos G., Goode T., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N., Hagopian D., Hagos B., Hall J., Healy C., Hegarty R., Honan T., Horn A., Houde N., Hughes L., Hunnicutt L., Husby M., Jester B., Jones C., Kamat A., Kanga B., Kells C., Khazanovich D., Kieu A. C., Kisner P., Kumar M., Lance K., Landers T., Lara M., Lee W., Leger J.-P., Lennon N., Leuper L., LeVine S., Liu J., Liu X., Lokyitsang Y., Lokyitsang T., Lui A., Macdonald J., Major J., Marabella R., Maru K., Matthews C., McDonough S., Mehta T., Meldrim J., Melnikov A., Meneus L., Mihalev A., Mihova T., Miller K., Mittelman R., Mlenga V., Mulrain L., Munson G., Navidi A., Naylor J., Nguyen T., Nguyen N., Nguyen C., Nguyen T., Nicol R., Norbu N., Norbu C., Novod N., Nyima T., Olandt P., O'Neill B., O'Neill K., Osman S., Oyono L., Patti C., Perrin D., Phunkhang P., Pierre F., Priest M., Rachupka A., Raghuraman S., Rameau R., Ray V., Raymond C., Rege F., Rise C., Rogers J., Rogov P., Sahalie J., Settipalli S., Sharpe T., Shea T., Sheehan M., Sherpa N., Shi J., Shih D., Sloan J., Smith C., Sparrow T., Stalker J., Stange-Thomann N., Stavropoulos S., Stone C., Stone S., Sykes S., Tchuinga P., Tenzing P., Tesfaye S., Thoulutsang D., Thoulutsang Y., Topham K., Topping I., Tsamla T., Vassiliev H., Venkataraman V., Vo A., Wangchuk T., Wangdi T., Weiand M., Wilkinson J., Wilson A., Yadav S., Yang S., Yang X., Young G., Yu Q., Zainoun J., Zembek L., Zimmer A., Lander E. S., 2005 Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature **438**: 803–819.

Liu F., van der Lijn F., Schurmann C., Zhu G., Chakravarty M. M., Hysi P. G., Wollstein A., Lao O., de Bruijne M., Ikram M. A., van der Lugt A., Rivadeneira F., Uitterlinden A. G., Hofman A., Niessen W. J., Homuth G., de Zubicaray G., McMahon K. L., Thompson P. M., Daboul A., Puls R., Hegenscheid K., Bevan L., Pausova Z., Medland S. E., Montgomery G. W., Wright M. J., Wicking C., Boehringer S., Spector T. D., Paus T., Martin N. G., Biffar R., Kayser M., 2012 A

genome-wide association study identifies five loci influencing facial morphology in Europeans. PLoS Genetics **8**: e1002932.

Liu N.-C., Adams V. J., Kalmar L., Ladlow J. F., Sargan D. R., 2016 Whole-Body Barometric Plethysmography Characterizes Upper Airway Obstruction in 3 Brachycephalic Breeds of Dogs. J Vet Intern Med **30**: 853–865.

Liu N.-C., Sargan D. R., Adams V. J., Ladlow J. F., 2015 Characterisation of Brachycephalic Obstructive Airway Syndrome in French Bulldogs Using Whole-Body Barometric Plethysmography. PLoS ONE **10**: e0130741.

Liu P., Lu J., Cardoso W. V., Vaziri C., 2007 The SPARC-related factor SMOC-2 promotes growth factor-induced cyclin D1 expression and DNA synthesis via integrin-linked kinase. Mol Biol Cell **19**: 248–261.

Liu R., Sun Y., Zhao G., Wang F., Wu D., Zheng M., Chen J., Zhang L., Hu Y., Wen J., 2013 Genome-Wide Association Study Identifies Loci and Candidate Genes for Body Composition and Meat Quality Traits in Beijing-You Chickens. PLoS ONE **8**: e61172.

Liu W., Selever J., Murali D., Sun X., Brugger S. M., Ma L., Schwartz R. J., Maxson R., Furuta Y., Martin J. F., 2005 Threshold-specific requirements for Bmp4 in mandibular development. Dev Biol **283**: 282–293.

Liu Y. H., Kundu R., Wu L., Luo W., Ignelzi M. A., Snead M. L., Maxson R. E., 1995 Premature suture closure and ectopic cranial bone in mice expressing Msx2 transgenes in the developing skull. Proceedings of the National Academy of Sciences of the United States of America **92**: 6137–6141.

Liu Y., Stein E., Oliver T., Li Y., Brunken W. J., Koch M., Tessier-Lavigne M., Hogan B. L. M., 2004 Novel role for Netrins in regulating epithelial behavior during lung branching morphogenesis. Curr Biol **14**: 897–905.

Love M. I., Huber W., Anders S., 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15**: 550.

Lumsden A., Sprawson N., Graham A., 1991 Segmental origin and migration of neural crest cells in the hindbrain region of the chick embryo. Development **113**: 1281–1291.

Lyons L. A., Erdman C. A., Grahn R. A., Hamilton M. J., Carter M. J., Helps C. R., Alhaddad H., Gandolfi B., 2015 Aristaless-Like Homeobox protein 1 (ALX1) variant associated with craniofacial structure and frontonasal dysplasia in Burmese cats. Dev Biol **409**: 451–458.

López-Lera A., Garrido S., Roche O., López-Trascasa M., 2011 SERPING1 mutations in 59 families with hereditary angioedema. Mol Immunol **49**: 18–27.

Ma L., Golden S., Wu L., Maxson R., 1996 The molecular basis of Boston-type craniosynostosis: the Pro148-->His mutation in the N-terminal arm of the MSX2 homeodomain stabilizes DNA binding without altering nucleotide sequence preferences. Hum Mol Genet **5**: 1915–1920.

Maccoux L. J., Clements D. N., Salway F., Day P. J. R., 2007 Identification of new reference genes for the normalisation of canine osteoarthritic joint tissue transcripts from microarray data. BMC Mol Biol **8**: 62.

Maier S., Paulsson M., Hartmann U., 2008 The widely expressed extracellular matrix protein SMOC-2 promotes keratinocyte attachment and migration. Exp Cell Res **314**: 2477–2487.

Marchant T. W., Johnson E. J., McTeir L., Johnson C. I., Gow A., Liuti T., Kuehn D., Svenson K., Bermingham M. L., Drögemüller M., Nussbaumer M., Davey M. G., Argyle D. J., Powell R. M., Guilherme S., Lang J., Haar ter G., Leeb T., Schwarz T., Mellanby R. J., Clements D. N., Schoenebeck J. J., 2017 Canine Brachycephaly Is Associated with a Retrotransposon-Mediated Missplicing of SMOC2. Curr Biol **27**: 1573–1584.e6.

Marouli E., Graff M., Medina-Gomez C., Lo K. S., Wood A. R., Kjaer T. R., Fine R. S., Lu Y., Schurmann C., Highland H. M., Rüeger S., Thorleifsson G., Justice A. E., Lamparter D., Stirrups K. E., Turcot V., Young K. L., Winkler T. W., Esko T., Karaderi T., Locke A. E., Masca N. G. D., Ng M. C. Y., Mudgal P., Rivas M. A., Vedantam S., Mahajan A., Guo X., Abecasis G., Aben K. K., Adair L. S., Alam D. S., Albrecht E., Allin K. H., Allison M., Amouyel P., Appel E. V., Arveiler D., Asselbergs F. W., Auer P. L., Balkau B., Banas B., Bang L. E., Benn M., Bergmann S., Bielak L. F., Blüher M., Boeing H., Boerwinkle E., Böger C. A., Bonnycastle L. L., Bork-

Jensen J., Bots M. L., Bottinger E. P., Bowden D. W., Brandslund I., Breen G., Brilliant M. H., Broer L., Burt A. A., Butterworth A. S., Carey D. J., Caulfield M. J., Chambers J. C., Chasman D. I., Chen Y.-D. I., Chowdhury R., Christensen C., Chu A. Y., Cocca M., Collins F. S., Cook J. P., Corley J., Galbany J. C., Cox A. J., Cuellar-Partida G., Danesh J., Davies G., de Bakker P. I. W., de Borst G. J., de Denus S., de Groot M. C. H., de Mutsert R., Deary I. J., Dedoussis G., Demerath E. W., Hollander den A. I., Dennis J. G., Di Angelantonio E., Drenos F., Du M., Dunning A. M., Easton D. F., Ebeling T., Edwards T. L., Ellinor P. T., Elliott P., Evangelou E., Farmaki A.-E., Faul J. D., Feitosa M. F., Feng S., Ferrannini E., Ferrario M. M., Ferrieres J., Florez J. C., Ford I., Fornage M., Franks P. W., Frikke-Schmidt R., Galesloot T. E., Gan W., Gandin I., Gasparini P., Giedraitis V., Giri A., Girotto G., Gordon S. D., Gordon-Larsen P., Gorski M., Grarup N., Grove M. L., Gudnason V., Gustafsson S., Hansen T., Harris K. M., Harris T. B., Hattersley A. T., Hayward C., He L., Heid I. M., Heikkilä K., Helgeland Ø., Hernesniemi J., Hewitt A. W., Hocking L. J., Hollensted M., Holmen O. L., Hovingh G. K., Howson J. M. M., Hoyng C. B., Huang P. L., Hveem K., Ikram M. A., Ingelsson E., Jackson A. U., Jansson J.-H., Jarvik G. P., Jensen G. B., Jhun M. A., Jia Y., Jiang X., Johansson S., Jørgensen M. E., Jørgensen T., Jousilahti P., Jukema J. W., Kahali B., Kahn R. S., Kähönen M., Kamstrup P. R., Kanoni S., Kaprio J., Karaleftheri M., Kardia S. L. R., Karpe F., Kee F., Keeman R., Kiemeney L. A., Kitajima H., Kluivers K. B., Kocher T., Komulainen P., Kontto J., Kooner J. S., Kooperberg C., Kovacs P., Kriebel J., Kuivaniemi H., Küry S., Kuusisto J., La Bianca M., Laakso M., Lakka T. A., Lange E. M., Lange L. A., Langefeld C. D., Langenberg C., Larson E. B., Lee I.-T., Lehtimäki T., Lewis C. E., Li H., Li J., Li-Gao R., Lin H., Lin L.-A., Lin X., Lind L., Lindström J., Linneberg A., Liu Y., Liu Y., Lophatananon A., Luan J., Lubitz S. A., Lyytikäinen L.-P., Mackey D. A., Madden P. A. F., Manning A. K., Männistö S., Marenne G., Marten J., Martin N. G., Mazul A. L., Meidtner K., Metspalu A., Mitchell P., Mohlke K. L., Mook-Kanamori D. O., Morgan A., Morris A. D., Morris A. P., Müller-Nurasyid M., Munroe P. B., Nalls M. A., Nauck M., Nelson C. P., Neville M., Nielsen S. F., Nikus K., Njølstad P. R., Nordestgaard B. G., Ntalla I., O'Connel J. R., Oksa H., Loohuis L.

M. O., Ophoff R. A., Owen K. R., Packard C. J., Padmanabhan S., Palmer C. N. A., Pasterkamp G., Patel A. P., Pattie A., Pedersen O., Peissig P. L., Peloso G. M., Pennell C. E., Perola M., Perry J. A., Perry J. R. B., Person T. N., Pirie A., Polasek O., Posthuma D., Raitakari O. T., Rasheed A., Rauramaa R., Reilly D. F., Reiner A. P., Renström F., Ridker P. M., Rioux J. D., Robertson N., Robino A., Rolandsson O., Rudan I., Ruth K. S., Saleheen D., Salomaa V., Samani N. J., Sandow K., Sapkota Y., Sattar N., Schmidt M. K., Schreiner P. J., Schulze M. B., Scott R. A., Segura-Lepe M. P., Shah S., Sim X., Sivapalaratnam S., Small K. S., Smith A. V., Smith J. A., Southam L., Spector T. D., Speliotes E. K., Starr J. M., Steinthorsdottir V., Stringham H. M., Stumvoll M., Surendran P., Hart L. M. '., Tansey K. E., Tardif J.-C., Taylor K. D., Teumer A., Thompson D. J., Thorsteinsdottir U., Thuesen B. H., Tönjes A., Tromp G., Trompet S., Tsafantakis E., Tuomilehto J., Tybjaerg-Hansen A., Tyrer J. P., Uher R., Uitterlinden A. G., Ulivi S., van der Laan S. W., Van Der Leij A. R., van Duijn C. M., van Schoor N. M., van Setten J., Varbo A., Varga T. V., Varma R., Edwards D. R. V., Vermeulen S. H., Vestergaard H., Vitart V., Vogt T. F., Vozzi D., Walker M., Wang F., Wang C. A., Wang S., Wang Y., Wareham N. J., Warren H. R., Wessel J., Willems S. M., Wilson J. G., Witte D. R., Woods M. O., Wu Y., Yaghootkar H., Yao J., Yao P., Yerges-Armstrong L. M., Young R., Zeggini E., Zhan X., Zhang W., Zhao J. H., Zhao W., Zhao W., Zheng H., Zhou W., Rotter J. I., Boehnke M., Kathiresan S., McCarthy M. I., Willer C. J., Stefansson K., Borecki I. B., Liu D. J., North K. E., Heard-Costa N. L., Pers T. H., Lindgren C. M., Oxvig C., Kutalik Z., Rivadeneira F., Loos R. J. F., Frayling T. M., Hirschhorn J. N., Deloukas P., Lettre G., 2017 Rare and low-frequency coding variants alter human adult height. Nature **542**: 186–190.

Masuda T., Yaginuma H., Sakuma C., Ono K., 2009 Netrin-1 signaling for sensory axons: Involvement in sensory axonal development and regeneration. Cell Adh Migr **3**: 171–173.

Matsushita M., Shichiri M., Imai T., Iwashina M., Tanaka H., Takasu N., Hirata Y., 2001 Co-expression of urotensin II and its receptor (GPR14) in human cardiovascular and renal tissues. J Hypertens **19**: 2185–2190.

Mauri N., Kleiter M., Dietschi E., Leschnik M., Högler S., Wiedmer M., Dietrich J., Henke D., Steffen F., Schuller S., Gurtner C., Stokar-Regenscheit N., O'Toole D., Bilzer T., Herden C., Oevermann A., Jagannathan V., Leeb T., 2017 A SINE Insertion in ATP1B2 in Belgian Shepherd Dogs Affected by Spongy Degeneration with Cerebellar Ataxia (SDCA2). G3 (Bethesda) **7**: 2729–2737.

Mayer U., Aumailley M., Mann K., Timpl R., Engel J., 1991 Calcium-dependent binding of basement membrane protein BM-40 (osteonectin, SPARC) to basement membrane collagen type IV. Eur J Biochem **198**: 141–150.

Mayousse V., Desquilbet L., Jeandel A., Blot S., 2017 Prevalence of neurological disorders in French bulldog: a retrospective study of 343 cases (2002-2016). BMC Vet Res **13**: 212.

McBratney-Owen B., Iseki S., Bamforth S. D., Olsen B. R., Morriss-Kay G. M., 2008 Development and tissue origins of the mammalian cranial base. Dev Biol **322**: 121–132.

McElroy M., Hayashi K., Garmy-Susini B., Kaushal S., Varner J. A., Moossa A. R., Hoffman R. M., Bouvet M., 2008 Fluorescent LYVE-1 antibody to image dynamically lymphatic trafficking of cancer cells in vivo. J Surg Res **151**: 68–73.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernysky 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.

Meischl C., Boer M., Ahlin A., Roos D., 2000 A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease. Eur J Hum Genet **8**: 697–703.

Melvin V. S., Feng W., Hernandez-Lagunas L., Artinger K. B., Williams T., 2013 A morpholino-based screen to identify novel genes involved in craniofacial morphogenesis. Dev Dyn **242**: 817–31.

Merlino F., Di Maro S., Yousif A. M., Caraglia M., Grieco P., 2013 Urotensin-II Ligands: An Overview from Peptide to Nonpeptide Structures. J Amino Acids **2013**: 979016.

Meulemans D., Bronner-Fraser M., 2002 Amphioxus and lamprey AP-2 genes: implications for neural crest evolution and migration patterns. Development **129**: 4953–4962.

Meyers G. A., Orlow S. J., Munro I. R., Przylepa K. A., Jabs E. W., 1995 Fibroblast growth factor receptor 3 (FGFR3) transmembrane mutation in Crouzon syndrome with acanthosis nigricans. Nat Genet **11**: 462–464.

Miki Y., Nishisho I., Horii A., Miyoshi Y., Utsunomiya J., Kinzler K. W., Vogelstein B., Nakamura Y., 1992 Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res **52**: 643–645.

Mitchell T. J., Knowler S. P., van den Berg H., Sykes J., Rusbridge C., 2014 Syringomyelia: determining risk and protective factors in the conformation of the Cavalier King Charles Spaniel dog. Canine Genet Epidemiol **1**: 9.

Mitteroecker P., Gunz P., 2009 Advances in Geometric Morphometrics. Evolutionary Biology **36**: 235–247.

Mommaerts H., Esguerra C. V., Hartmann U., Luyten F. P., Tylzanowski P., 2014 Smoc2 modulates embryonic myelopoiesis during zebrafish development. Dev Dyn **243**: 1375–1390.

Monsoro-Burq A. H., 2015 PAX transcription factors in neural crest development. Semin Cell Dev Biol **44**: 87–96.

Moore K. L., Persuad T. V. N. (Eds.), 2003 The Pharyngeal Apparatus. In: *Before we are Born Essentials of Embryology and Birth Defects*, Saunders Elsevier, pp. 151–188.

Moore M. H., 1993 Upper airway obstruction in the syndromal craniosynostoses. Br J Plast Surg **46**: 355–362.

Mount S. M., 1982, A catalogue of splice junction sequences. Nucleic Acids Res. **10:** 459-472.

Muenke M., Gripp K. W., McDonald-McGinn D. M., Gaudenz K., Whitaker L. A., Bartlett S. P., Markowitz R. I., Robin N. H., Nwokoro N., Mulvihill J. J., Losken H. W., Mulliken J. B., Guttmacher A. E., Wilroy R. S., Clarke L. A., Hollway G., Adès L. C., Haan E. A., Mulley J. C., Cohen M. M., Bellus G. A., Francomano C. A., Moloney

D. M., Wall S. A., Wilkie A. O., 1997 A unique point mutation in the fibroblast growth factor receptor 3 gene (FGFR3) defines a new craniosynostosis syndrome. Am J Hum Genet **60**: 555–564.

Muenke M., Schell U., Hehr A., Robin N. H., Losken H. W., Schinzel A., Pulleyn L. J., Rutland P., Reardon W., Malcolm S., 1994 A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome. Nat Genet **8**: 269–274.

Mullin S. K., Taylor P. J., 2002 The effects of parallax on geometric morphometric data. Comput Biol Med **32**: 455–464.

Murphy W. J., Eizirik E., Johnson W. E., Zhang Y. P., Ryder O. A., O'Brien S. J., 2001 Molecular phylogenetics and the origins of placental mammals. Nature **409**: 614–618.

Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O., 1997 Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. DNA Res **4**: 141–150.

Nakamura T., Takeuchi K., Muraoka S., Takezoe H., Takahashi N., Mori N., 1999 A neurally enriched coronin-like protein, ClipinC, is a novel candidate for an actin cytoskeleton-cortical membrane-linking protein. Journal of Biological Chemistry **274**: 13322–13327.

Nakata N., Yu P. I., Davis B., Nance W. E., 1973 The use of genetic data in the prediction of craniofacial dimensions. Am J Orthod **63**: 471–480.

Narita N., Nishio H., Kitoh Y., Ishikawa Y., Ishikawa Y., Minami R., Nakamura H., Matsuo M., 1993 Insertion of a 5“ truncated L1 element into the 3” end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. J Clin Invest **91**: 1862–1867.

Neff M. W., Beck J. S., Koeman J. M., Boguslawski E., Kefene L., Borgman A., Ruhe A. L., 2012 Partial deletion of the sulfate transporter SLC13A1 is associated with an osteochondrodysplasia in the Miniature Poodle breed. PLoS ONE **7**: e51917.

Nichols D. H., 1981 Neural crest formation in the head of the mouse embryo as observed using a new histological technique. J Embryol Exp Morphol **64**: 105–120.

Nishimura R., Hata K., Ono K., Amano K., Takigawa Y., Wakabayashi M., Takashima R., Yoneda T., 2012 Regulation of endochondral ossification by transcription factors. Front Biosci (Landmark Ed) **17**: 2657–2666.

Nixon G. M., Brouillette R. T., 2005 Sleep . 8: paediatric obstructive sleep apnoea. Thorax **60**: 511–516.

Noden D. M., 1978 The control of avian cephalic neural crest cytodifferentiation. I. Skeletal and connective tissues. Dev Biol **67**: 296–312.

Noden D. M., Trainor P. A., 2005 Relations and interactions between cranial mesoderm and neural crest populations. J Anat **207**: 575–601.

Nussbaumer M., 1978 Biometric Analysis of the Skull Base in Small and Medium Sized Dogs. Tierzüchtung Züchtung Biol: 1–14.

O'Neill D. G., Jackson C., Guy J. H., Church D. B., McGreevy P. D., Thomson P. C., Brodbelt D. C., 2015 Epidemiological associations between brachycephaly and upper respiratory tract disorders in dogs attending veterinary practices in England. Canine Genet Epidemiol **2**: 10.

O'Neill D. G., Lee M. M., Brodbelt D. C., Church D. B., Sanchez R. F., 2017 Corneal ulcerative disease in dogs under primary veterinary care in England: epidemiology and clinical management. Canine Genet Epidemiol **4**: 5.

O'Neill D.G., O'Sullivan A.M., Manson E.A., Church D.B., Goag A.K., McGreevy P.D., Brodbelt D.C., 2017 Canine dystocia in 50 UK first-opinion emergency-care veterinary practices: prevalence and risk factors. **181**: 88.

O'Rahilly R., Müller F., 2007 The development of the neural crest in the human. J Anat **211**: 335–351.

Oberlender S. A., Tuan R. S., 1994 Expression and functional involvement of N-cadherin in embryonic limb chondrogenesis. Development **120**: 177–187.

Oechtering G. U., Pohl S., Schlueter C., Schuenemann R., 2016 A Novel Approach to Brachycephalic Syndrome. 2. Laser-Assisted Turbinectomy (LATE). Vet Surg **45**: 173–181.

Office for National Statistics, 2018 Consumer trends, UK: January to March 2018. https://wwwonsgovuk/economy/nationalaccounts/satelliteaccounts/bulletins/consumertrends/januarytomarch2018.

Ogino H., Hisanaga A., Kohno T., Kondo Y., Okumura K., Kamei T., Sato T., Asahara H., Tsuiji H., Fukata M., Hattori M., 2017 Secreted Metalloproteinase ADAMTS-3 Inactivates Reelin. J Neurosci **37**: 3181–3191.

Olderøy M. Ø., Lilledahl M. B., Beckwith M. S., Melvik J. E., Reinholt F., Sikorski P., Brinchmann J. E., 2014 Biochemical and structural characterization of neocartilage formed by mesenchymal stem cells in alginate hydrogels. PLoS ONE **9**: e91662.

Oldridge M., Zackai E. H., McDonald-McGinn D. M., Iseki S., Morriss-Kay G. M., Twigg S. R., Johnson D., Wall S. A., Jiang W., Theda C., Jabs E. W., Wilkie A. O., 1999 De novo alu-element insertions in FGFR2 identify a distinct pathological basis for Apert syndrome. Am J Hum Genet **64**: 446–461.

Ollmann M. M., Wilson B. D., Yang Y. K., Kerns J. A., Chen Y., Gantz I., Barsh G. S., 1997 Antagonism of central melanocortin receptors in vitro and in vivo by agouti-related protein. Science **278**: 135–138.

Ostrander E. A., Kruglyak L., 2000 Unleashing the canine genome. Genome Res **10**: 1271–1274.

Ostrander E. A., Galibert F., Patterson D. F., 2000 Canine genetics comes of age. Trends Genet **16**: 117–124.

Otto A. W., Starck D., Starck D., 1831 *A Compendium of Human and Comparative Pathological Anatomy*. London.

Ovodov N. D., Crockford S. J., Kuzmin Y. V., Higham T. F. G., Hodgins G. W. L., 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.

Pazin D.E., Albrecht K.H., 2009 Developmental expression of Smoc1 and Smoc2 suggests potential roles in fetal gonad and reproductive tract differentiation. Dev Dyn **238**: 2877.

PFMA, 2018 PFMA's Pet Data Report 2018. https://wwwpfmaorguk/_assets/docs/annual-reports/PFMA-Pet-Data-Report-2018pdf.

Packer R. M. A., A H., CC B., 2012 Do dog owners perceive the clinical signs related to conformational inherited disorders as “normal” for the breed? A potential constraint to improving canine welfare. Animal Welfare **21**: 81–93.

Packer R. M. A., Hendricks A., Burn C. C., 2015 Impact of facial conformation on canine health: corneal ulceration. PLoS ONE **10**: e0123827.

Parker H. G., Kim L. V., Sutter N. B., Carlson S., Lorentzen T. D., Malek T. B., Johnson G. S., DeFrance H. B., Ostrander E. A., Kruglyak L., 2004 Genetic structure of the purebred domestic dog. Science **304**: 1160–1164.

Parker H. G., vonHoldt B. M., Quignon P., Margulies E. H., Shao S., Mosher D. S., Spady T. C., Elkahloun A., Cargill M., Jones P. G., Maslen C. L., Acland G. M., Sutter N. B., Kuroki K., Bustamante C. D., Wayne R. K., Ostrander E. A., 2009 An expressed fgf4 retrogene is associated with breed-defining chondrodysplasia in domestic dogs. Science **325**: 995–998.

Patel S. R., Goodloe R., De G., Kowgier M., Weng J., Buxbaum S. G., Cade B., Fulop T., Gharib S. A., Gottlieb D. J., Hillman D., Larkin E. K., Lauderdale D. S., Li L., Mukherjee S., Palmer L., Zee P., Zhu X., Redline S., 2012 Association of genetic loci with sleep apnea in European Americans and African-Americans: the Candidate Gene Association Resource (CARe). PLoS ONE **7**: e48836.

Paternoster L., Zhurov A. I., Toma A. M., Kemp J. P., Pourcain B. S., Timpson N. J., McMahon G., McArdle W., Ring S. M., Smith G. D., Richmond S., Evans D. M., 2012 Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. Am J Hum Genet **90**: 478–485.

Patterson D. F., 2000 Companion animal medicine in the age of medical genetics. J Vet Intern Med **14**: 1–9.

Pearson K., 1901 On Lines and Planes of Closest Fit to Systems of Points in Space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **2**: 559–572.

Pedersen N. C., Pooch A. S., Liu H., 2016 A genetic assessment of the English bulldog. Canine Genet Epidemiol **3**: 6.

Peeters T., Monteagudo S., Tylzanowski P., Luyten F. P., Lories R., Cailotto F., 2018 SMOC2 inhibits calcification of osteoprogenitor and endothelial cells. PLoS ONE **13**: e0198104.

Perepelitsa-Belancio V., Deininger P., 2003 RNA truncation by premature polyadenylation attenuates human mobile element activity. Nat Genet **35**: 363–366.

Perris R., 1997 The extracellular matrix in neural crest-cell migration. Trends Neurosci **20**: 23–31.

Perris R., Perissinotto D., 2000 Role of the extracellular matrix during neural crest cell migration. Mech Dev **95**: 3–21.

Persing J. A., 2008 MOC-PS(SM) CME article: management considerations in the treatment of craniosynostosis. Plast Reconstr Surg **121**: 1–11.

Persing J. A., Jane J. A., Shaffrey M., 1989 Virchow and the pathogenesis of craniosynostosis: a translation of his original work. Plast Reconstr Surg **83**: 738–742.

Pfaffl M.W., 2001, A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res **29**: 45e.

Pihlajamäki J., Lerin C., Itkonen P., Boes T., Floss T., Schroeder J., Dearie F., Crunkhorn S., Burak F., Jimenez-Chillaron J. C., Kuulasmaa T., Miettinen P., Park P. J., Nasser I., Zhao Z., Zhang Z., Xu Y., Wurst W., Ren H., Morris A. J., Stamm S., Goldfine A. B., Laakso M., Patti M. E., 2011 Expression of the splicing factor gene SFRS10 is reduced in human obesity and contributes to enhanced lipogenesis. Cell Metab **14**: 208–218.

Pingault V., Ente D., Moal F. D.-L., Goossens M., Marlin S., Bondurand N., 2010 Review and update of mutations causing Waardenburg syndrome. Hum Mutat **31**: 391–406.

Pionnier-Capitan M., Bemilli C., Bodu P., Célérier G., Ferrié J.-G., Fosse P., Garcià M., Vigne J.-D., 2011 New evidence for Upper Palaeolithic small domestic dogs in South-Western Europe. Journal of Archaeological Science **38**: 2123.

Plassais J., Rimbault M., Williams F. J., Davis B. W., Schoenebeck J. J., Ostrander E. A., 2017 Analysis of large versus small dogs reveals three genes on the canine X chromosome associated with body weight, muscling and back fat thickness. PLoS Genetics **13**: e1006661.

Poncet C. M., Dupre G. P., Freiche V. G., Estrada M. M., Poubanne Y. A., Bouvy B. M., 2005a Prevalence of gastrointestinal tract lesions in 73 brachycephalic dogs with upper respiratory syndrome. J Small Anim Pract **46**: 273–279.

Poncet C. M., Dupre G. P., Freiche V. G., Estrada M. M., Poubanne Y. A., Bouvy B. M., 2005b Prevalence of gastrointestinal tract lesions in 73 brachycephalic dogs with upper respiratory syndrome. J Small Anim Pract **46**: 273–279.

Poulos M. G., Batra R., Li M., Yuan Y., Zhang C., Darnell R. B., Swanson M. S., 2013 Progressive impairment of muscle regeneration in muscleblind-like 3 isoform knockout mice. Hum Mol Genet **22**: 3547–3558.

Pretzer S. D., 2008 Canine embryonic and fetal development: a review. Theriogenology **70**: 300–303.

Pritchard J. K., Stephens M., Donnelly P., 2000 Inference of population structure using multilocus genotype data. Molecular Ecology Notes **155**: 945–959.

Pruitt KD., Tatusova T., Maglott DR., 2005 NCBI Reference Sequence (ReSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucelic Acids Res. **33**: D501-D504.

Pruitt KD., Brown GR>< Hiatt SM., Thibaud-Nissen F., Astashyn A., Ermolaeva O., Farrel CM., Hart J., Landrum MJ., McGarvey KM., Murphy MR., O'Leary NA., Pujar S., Rajput B., Rangwala SH., Riddick LD., Shkeda A., Sun H., Tamez P., Tully RE., Wallin C., Webb D., Weber J., Wu W., Dicuccio M., Kitts P., Maglott DR., Murphy

TD.,Ostell JM., 2014 RefSeq: an update on mammalian reference sequences. Nucelic Acids Res. **42**: D756-D763.

Pugh C. A., de C Bronsvoort B. M., Handel I. G., Summers K. M., Clements D. N., 2015 Dogslife: A cohort study of Labrador Retrievers in the UK. Prev Vet Med **122**: 426–435.

Pulichino A.-M., Vallette-Kasic S., Couture C., Gauthier Y., Brue T., David M., Malpuech G., Deal C., Van Vliet G., De Vroede M., Riepe F. G., Partsch C.-J., Sippell W. G., Berberoglu M., Atasay B., Drouin J., 2003 Human and mouse TPIT gene mutations cause early onset pituitary ACTH deficiency. Genes Dev **17**: 711–716.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A. R., Bender D., Maller J., Sklar P., de Bakker P. I. W., 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**: 559–575.

Raj A., Stephens M., PRITCHARD J. K., 2014 fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics **197**: 573–589.

Ran F. A., Cox D., Lin S., Barretto R., Habib N., Hsu P. D., Wu X., Jiang W., Marraffini L. A., Zhang F., 2013 Multiplex genome engineering using CRISPR/Cas systems. Science **339**: 819–823.

Reardon W., Winter R. M., Rutland P., Pulleyn L. J., Jones B. M., Malcolm S., 1994 Mutations in the fibroblast growth factor receptor 2 gene cause Crouzon syndrome. Nat Genet **8**: 98–103.

Regodón S., Vivo J. M., Franco A., Guillén M. T., Robina A., 1993 Craniofacial angle in dolicho-, meso- and brachycephalic dogs: radiological determination and application. Annals of Anatomy - Anatomischer Anzeiger **175**: 361.

Riecks T. W., Birchard S. J., Stephens J. A., 2007 Surgical correction of brachycephalic syndrome in dogs: 62 cases (1991–2004). J Am Vet Med Assoc **230**: 1324.

Rieder S., Taourit S., Mariat D., Langlois B., Guérin G., 2001 Mutations in the agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (Equus caballus). Mamm Genome **12**: 450–455.

Rigueur D., Lyons K. M., 2014 Whole-mount skeletal staining. Methods Mol Biol **1130**: 113–121.

Rimbault M., Beale H. C., Schoenebeck J. J., Hoopes B. C., Allen J. J., Kilroy-Glynn P., Wayne R. K., Sutter N. B., Ostrander E. A., 2013 Derived variants at six genes explain nearly half of size reduction in dog breeds. Genome Res **23**: 1985–1995.

Risch N., Merikangas K., 1996 The future of genetic studies of complex human diseases. Science **273**: 1516–1517.

Robinson J. T., Thorvaldsdóttir H., Winckler W., Guttman M., Lander E. S., Getz G., Mesirov J. P., 2011 Integrative genomics viewer. Nat Biotechnol **29**: 24–26.

Rocnik E. F., Liu P., Sato K., Walsh K., Vaziri C., 2006 The novel SPARC family member SMOC-2 potentiates angiogenic growth factor activity. Journal of Biological Chemistry **281**: 22855–22864.

Rogg M., Yasuda-Yamahara M., Abed A., Dinse P., Helmstädter M., Conzelmann A. C., Frimmel J., Sellung D., Biniossek M. L., Kretz O., Grahammer F., Schilling O., Huber T. B., Schell C., 2017 The WD40-domain containing protein CORO2B is specifically enriched in glomerular podocytes and regulates the ventral actin cytoskeleton. Sci Rep **7**: 15910.

Rohlf F. J., Marcus L. F., 1993 A revolution morphometrics. Trends Ecol Evol (Amst) **8**: 129–132.

Roshyara N. R., Scholz M., 2014 fcGENE: a versatile tool for processing and transforming SNP datasets. PLoS ONE **9**: e97589.

Rutkovskiy A., Stensløkken K.-O., Vaage I. J., 2016 Osteoblast Differentiation at a Glance. Med Sci Monit Basic Res **22**: 95–106.

Rutland P., Pulleyn L. J., Reardon W., Baraitser M., Hayward R., Jones B., Malcolm S., Winter R. M., Oldridge M., Slaney S. F., 1995 Identical mutations in the FGFR2 gene cause both Pfeiffer and Crouzon syndrome phenotypes. Nat Genet **9**: 173–176.

Ryan R., Gutierrez-Quintana R., Haar Ter G., De Decker S., 2017 Prevalence of thoracic vertebral malformations in French bulldogs, Pugs and English bulldogs with and without associated neurological deficits. Vet J **221**: 25–29.

Sahana G., Höglund J. K., Guldbrandtsen B., Lund M. S., 2015 Loci associated with adult stature also affect calf birth survival in cattle. BMC Genet **16**: 47.

Salminen M., Meyer B. I., Bober E., Gruss P., 2000 Netrin 1 is required for semicircular canal formation in the mouse inner ear. Development **127**: 13–22.

Sanchez R. F., Innocent G., Mould J., Billson F. M., 2007 Canine keratoconjunctivitis sicca: disease trends in a review of 229 cases. J Small Anim Pract **48**: 211–217.

Sarimski K., 1998 Children with Apert syndrome: behavioural problems and family stress. Dev Med Child Neurol **40**: 44–49.

Sarimski K., 2001 Social adjustment of children with a severe craniofacial anomaly (Apert syndrome). Child Care Health Dev **27**: 583–590.

Sasaki T., Göhring W., Mann K., Maurer P., Hohenester E., Knäuper V., Murphy G., Timpl R., 1997 Limited cleavage of extracellular matrix protein BM-40 by matrix metalloproteinases increases its affinity for collagens. Journal of Biological Chemistry **272**: 9237–9243.

Sasaki T., Hohenester E., Göhring W., Timpl R., 1998 Crystal structure and mapping by site-directed mutagenesis of the collagen-binding epitope of an activated form of BM-40/SPARC/osteonectin. EMBO J **17**: 1625–1634.

Savolainen P., Zhang Y.-P., Luo J., Lundeberg J., Leitner T., 2002 Genetic evidence for an East Asian origin of domestic dogs. Science **298**: 1610–1613.

Savoye I., Loos R., Carels C., Derom C., Vlietinck R., 1998 A genetic study of anteroposterior and vertical facial proportions using model-fitting. Angle Orthod **68**: 467–470.

Schlueter C., Budras K. D., Ludewig E., Mayrhofer E., Koenig H. E., Walter A., Oechtering G. U., 2009 Brachycephalic feline noses: CT and anatomical study of the relationship between head conformation and the nasolacrimal drainage system. J Feline Med Surg **11**: 891–900.

Schmahl J., Raymond C. S., Soriano P., 2006 PDGF signaling specificity is mediated through multiple immediate early genes. Nat Genet **39**: 52–60.

Schneider C. A., Rasband W. S., Eliceiri K. W., 2012 NIH Image to ImageJ: 25 years of image analysis. Nat Methods **9**: 671–675.

Schoenwolf G. C., Bleyl S. B., Brauer P. R., Francis-West P. H. (Eds.), 2015 Development of the Pharyngeal Apparatus and Face. In: *Larsen's Human Embryology*, Elsevier Saunders, pp. 429–472.

Schroeder T. M., Westendorf J. J., 2005 Histone deacetylase inhibitors promote osteoblast maturation. J Bone Miner Res **20**: 2254–2263.

Schoenebeck J.J., Hutchinson S.A., Byers A., Beale H.C., Carrington B., Faden D.L., Rimbault M., Decker B., Kidd J.M., Sood R., Boyko A.R., Fondon J.W., Wayne R.K., Bustamante C.D., Ciruna B., Ostrander E.A., 2012 Variation of BMP3 Contributes to Dog Breed Skull Diversity. PLoS Genetics **8**: e1002849.

Schulz A., Jundt G., Berghäuser K. H., Gehron-Robey P., Termine J. D., 1988 Immunohistochemical study of osteonectin in various types of osteosarcoma. Am J Pathol **132**: 233–238.

Schwahn U., Lenzner S., Dong J., Feil S., Hinzmann B., van Duijnhoven G., Kirschner R., Hemberger M., Bergen A. A., Rosenberg T., Pinckers A. J., Fundele R., Rosenthal A., Cremers F. P., Ropers H. H., Berger W., 1998 Positional cloning of the gene for X-linked retinitis pigmentosa 2. Nat Genet **19**: 327–332.

Scott C. K., Hightower J. A., 1991 The matrix of endochondral bone differs from the matrix of intramembranous bone. Calcif Tissue Int **49**: 349–354.

Sears K. E., Goswami A., Flynn J. J., Niswander L. A., 2007 The correlated evolution of Runx2 tandem repeats, transcriptional activity, and facial length in carnivora. Evol Dev **9**: 555–565.

Shearin A. L., Ostrander E. A., 2010 Leading the way: canine models of genomics and disease. Dis Model Mech **3**: 27–34.

Shingleton A. W., Frankino W. A., 2012 New perspectives on the evolution of exaggerated traits. Bioessays **35**: 100–107.

Shingleton A. W., Estep C. M., Driscoll M. V., Dworkin I., 2009 Many ways to be small: different environmental regulators of size generate distinct scaling relationships in Drosophila melanogaster. Proc Biol Sci **276**: 2625–2633.

Sin D. D., Fitzgerald F., Parker J. D., Newton G., Floras J. S., Bradley T. D., 1999 Risk factors for central and obstructive sleep apnea in 450 men and women with congestive heart failure. Am J Respir Crit Care Med **160**: 1101–1106.

Slavotinek A. M., Baranzini S. E., Schanze D., Labelle-Dumais C., Short K. M., Chao R., Yahyavi M., Bijlsma E. K., Chu C., Musone S., Wheatley A., Kwok P.-Y., Marles S., Fryns J.-P., Maga A. M., Hassan M. G., Gould D. B., Madireddy L., Li C., Cox T. C., Smyth I., Chudley A. E., Zenker M., 2011 Manitoba-oculo-tricho-anal (MOTA) syndrome is caused by mutations in FREM1. J Med Genet **48**: 375–382.

Smith B. F., Yue Y., Woods P. R., Kornegay J. N., Shin J.-H., Williams R. R., Duan D., 2010 An intronic LINE-1 element insertion in the dystrophin gene aborts dystrophin expression and results in Duchenne-like muscular dystrophy in the corgi breed. Lab Invest **91**: 216–231.

Smith B. G., Hutcheson K. A., Little L. G., Skoracki R. J., Rosenthal D. I., Lai S. Y., Lewin J. S., 2014 Lymphedema outcomes in patients with head and neck cancer. Otolaryngol Head Neck Surg **152**: 284–291.

Song C., Gu X., Feng C., Wang Y., Gao Y., Hu X., Li N., 2010 Evaluation of SNPs in the chicken HMGA2 gene as markers for body weight gain. Animal Genetics **42**: 333–336.

Speek M., 2001 Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol **21**: 1973–1985.

Speltz M. L., Collett B. R., Wallace E. R., Starr J. R., Cradock M. M., Buono L., Cunningham M., Kapp-Simon K., 2015 Intellectual and academic functioning of school-age children with single-suture craniosynostosis. Pediatrics **135**: e615–23.

St-Jacques B., Hammerschmidt M., McMahon A. P., 1999 Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. Genes Dev **13**: 2072–2086.

Stein G. S., Lian J. B., Gerstenfeld L. G., Shalhoub V., Aronow M., Owen T., Markose E., 1989 The onset and progression of osteoblast differentiation is functionally related to cellular proliferation. Connect Tissue Res **20**: 3–13.

Stockard C. R., 1941 *The Genetic and Endocrinic Basis for Differences in Form and Behaviour*. Philadelphia: The Wistar Institute of Anatomy and Biology.

Sun Y., Huang Y., Yin A., Pan Y., Wang Y., Wang C., Du Y., Wang M., Lan F., Hu Z., Wang G., Jiang M., Ma J., Zhang X., Ma H., Ma J., Zhang W., Huang Q., Zhou Z., Ma L., Li Y., Jiang H., Xie L., Jiang Y., Shi B., Cheng J., Shen H., Wang L., Yang Y., 2015 Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. Nat Commun **6**: 6414.

Suto J.-I., 2008 Identification of multiple quantitative trait loci affecting the size and shape of the mandible in mice. Mamm Genome **20**: 1–13.

Suto J.-I., 2009 The Ay allele at the agouti locus reduces the size and alters the shape of the mandible in mice. Proc Jpn Acad, Ser B, Phys Biol Sci **85**: 248–257.

Sutter N. B., Bustamante C. D., Chase K., Gray M. M., Zhao K., Zhu L., Padhukasahasram B., Karlins E., Davis S., Jones P. G., Quignon P., Johnson G. S., Parker H. G., Fretwell N., Mosher D. S., Lawler D. F., Satyaraj E., Nordborg M., Lark K. G., Wayne R. K., Ostrander E. A., 2007 A single IGF1 allele is a major determinant of small size in dogs. Science **316**: 112–115.

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**: 2388–2396.

Sutter N. B., Mosher D. S., Gray M. M., Ostrander E. A., 2008 Morphometrics within dog breeds are highly reproducible and dispute Rensch's rule. Mamm Genome **19**: 713–723.

Symer D. E., Connelly C., Szak S. T., Caputo E. M., Cost G. J., Parmigiani G., Boeke J. D., 2002 Human l1 retrotransposition is associated with genetic instability in vivo. Cell **110**: 327–338.

Szak S. T., Pickeral O. K., Makalowski W., Boguski M. S., Landsman D., Boeke J. D., 2002 Molecular archeology of L1 insertions in the human genome. Genome Biol **3**: research0052.

Tabor H. K., Risch N. J., Myers R. M., 2002 Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet **3**: 391–397.

Takarada T., Nakazato R., Tsuchikane A., Fujikawa K., Iezaki T., Yoneda Y., Hinoi E., 2015 Genetic analysis of Runx2 function during intramembranous ossification. Development **143**: 211–218.

Tamburrini G., Caldarellia M., Massimi L., Santini P., Di Rocco C., 2005 Intracranial pressure monitoring in children with single suture and complex craniosynostosis: a review. Childs Nerv Syst. **10**: 913-921

Tchernov E., Valla F. F., 1997 Two New Dogs, and Other Natufian Dogs, from the Southern Levant. Journal of Archaeological Science **24**: 65.

Termine J. D., Kleinman H. K., Whitson S. W., Conn K. M., McGarvey M. L., Martin G. R., 1981 Osteonectin, a bone-specific protein linking mineral to collagen. Cell **26**: 99–105.

Tetens J., Widmann P., Kühn C., Thaller G., 2013 A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. Animal Genetics **44**: 467–471.

Thalmann O., Shapiro B., Cui P., Schuenemann V. J., Sawyer S. K., Greenfield D. L., Germonpré M. B., Sablin M. V., López-Giráldez F., Domingo-Roura X., Napierala H., Uerpmann H.-P., Loponte D. M., Acosta A. A., Giemsch L., Schmitz R. W., Worthington B., Buikstra J. E., Druzhkova A., Graphodatsky A. S., Ovodov N. D., Wahlberg N., Freedman A. H., Schweizer R. M., Koepfli K.-P., Leonard J. A., Meyer M., Krause J., Pääbo S., Green R. E., Wayne R. K., 2013 Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. Science **342**: 871–874.

The Kennel Club., 2018a *The New Complete Dog Book*.

The Kennel Club., 2018b Summary of Registrations for All Groups. https://wwwthekennelcluborguk/media/128966/group_statspdf.

Thorleifsson G., Walters G. B., Gudbjartsson D. F., Steinthorsdottir V., Sulem P., Helgadottir A., Gretarsdottir S., Thorlacius S., Jonsdottir I., Jonsdottir T., Olafsdottir E. J., Olafsdottir G. H., Jonsson T., Jonsson F., Borch-Johnsen K., Hansen T., Andersen G., Lauritzen T., Aben K. K., Verbeek A. L. M., Roeleveld N., Kampman E., Yanek L. R., Becker L. C., Tryggvadottir L., Rafnar T., Becker D. M., Gulcher J., Kiemeney L. A., Pedersen O., Kong A., Thorsteinsdottir U., Stefansson K., 2008 Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet **41**: 18–24.

Torrez C. V., Hunt G. B., 2006 Results of surgical correction of abnormalities associated with brachycephalic airway obstruction syndrome in dogs in Australia. J Small Anim Pract **47**: 150–154.

Trainor P. A., Sobieszczuk D., Wilkinson D., Krumlauf R., 2002 Signalling between the hindbrain and paraxial tissues dictates neural crest migration pathways. Development **129**: 433–442.

Trut L., Oskina I., Kharlamova A., 2009 Animal evolution during domestication: the domesticated fox as a model. Bioessays **31**: 349–360.

Twigg S. R. F., Kan R., Babbs C., Bochukova E. G., Robertson S. P., Wall S. A., Morriss-Kay G. M., Wilkie A. O. M., 2004 Mutations of ephrin-B1 (EFNB1), a marker of tissue boundary formation, cause craniofrontonasal syndrome. Proc Natl Acad Sci USA **101**: 8652–8657.

Twigg S. R. F., Versnel S. L., Nürnberg G., Lees M. M., Bhat M., Hammond P., Hennekam R. C. M., Hoogeboom A. J. M., Hurst J. A., Johnson D., Robinson A. A., Scambler P. J., Gerrelli D., Nürnberg P., Mathijssen I. M. J., Wilkie A. O. M., 2009 Frontorhiny, a distinctive presentation of frontonasal dysplasia caused by recessive mutations in the ALX3 homeobox gene. Am J Hum Genet **84**: 698–705.

Uhlén M., Fagerberg L., Hallström B. M., Lindskog C., Oksvold P., Mardinoglu A., Sivertsson Å., Kampf C., Sjöstedt E., Asplund A., Olsson I., Edlund K., Lundberg E., Navani S., Szigyarto C. A.-K., Odeberg J., Djureinovic D., Takanen J. O., Hober S., Alm T., Edqvist P.-H., Berling H., Tegel H., Mulder J., Rockberg J., Nilsson P., Schwenk J. M., Hamsten M., Feilitzen von K., Forsberg M., Persson L., Johansson

F., Zwahlen M., Heijne von G., Nielsen J., Pontén F., 2015 Proteomics. Tissue-based map of the human proteome. Science **347**: 1260419.

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B. C., Remm M., Rozen S. G., 2012 Primer3--new capabilities and interfaces. Nucleic Acids Res **40**: e115.

Van der Auwera G. A., Carneiro M. O., Hartl C., Poplin R., del Angel G., Levy-Moonshine A., Jordan T., Shakir K., Roazen D., Thibault J., Banks E., Garimella K. V., Altshuler D., Gabriel S., DePristo M. A., 2014 From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics **43**: 11.10.1–33.

van Steenbeek F. G., Spee B., Penning L. C., Kummeling A., van Gils I. H. M., Grinwis G. C. M., Van Leenen D., Holstege F. C. P., Vos-Loohuis M., Rothuizen J., Leegwater P. A. J., 2013 Altered subcellular localization of heat shock protein 90 is associated with impaired expression of the aryl hydrocarbon receptor pathway in dogs. PLoS ONE **8**: e57973

Vannahme C., Gösling S., Paulsson M., Maurer P., Hartmann U., 2003 Characterization of SMOC-2, a modular extracellular calcium-binding protein. Biochem J **373**: 805–14.

Vaysse A., Ratnakumar A., Derrien T., Axelsson E., Pielberg G. R., Sigurdsson S., Fall T., Bannasch D., Vilà C., Lohi H., Galibert F., Fredholm M., Häggström J., André C., Lindblad-Toh K., Hitte C., 2011 Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping. **7**: e1002316.

Vega H., Waisfisz Q., Gordillo M., Sakai N., Yanagihara I., Yamada M., van Gosliga D., Kayserili H., Xu C., Ozono K., Jabs E. W., Inui K., Joenje H., 2005 Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. Nat Genet **37**: 468–470.

Veitia R. A., 2007 Exploring the molecular etiology of dominant-negative mutations. Plant Cell **19**: 3843–3851.

Veitschegger K., Wilson L. A. B., Nussberger B., Camenisch G., Keller L. F., Wroe S., Sánchez-Villagra M. R., 2018 Resurrecting Darwin's Niata - anatomical, biomechanical, genetic, and morphometric studies of morphological novelty in cattle. Sci Rep **8**: 9129.

Velázquez-Aragón J. A., Alcántara-Ortigoza M. A., Estandia-Ortega B., Reyna-Fabián M. E., Méndez-Adame C. D., Angel A. G.-D., 2016 Gene Interactions Provide Evidence for Signaling Pathways Involved in Cleft Lip/Palate in Humans. J Dent Res **95**: 1257–1264.

Vondenhoff M. F., van de Pavert S. A., Dillard M. E., Greuter M., Goverse G., Oliver G., Mebius R. E., 2008 Lymph sacs are not required for the initiation of lymph node formation. Development **136**: 29–34.

Vrieling H., Duhl D. M., Millar S. E., Miller K. A., Barsh G. S., 1994 Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene. Proceedings of the National Academy of Sciences of the United States of America **91**: 5667–5671.

Wang W., Kirkness E. F., 2005 Short interspersed elements (SINEs) are a major source of canine genomic diversity. Genome Res **15**: 1798–1808.

Wang Y., Xiao R., Yang F., Karim B. O., Iacovelli A. J., Cai J., Lerner C. P., Richtsmeier J. T., Leszl J. M., Hill C. A., Yu K., Ornitz D. M., Elisseeff J., Huso D. L., Jabs E. W., 2005 Abnormalities in cartilage and bone development in the Apert syndrome FGFR2(+/S252W) mouse. Development **132**: 3537–3548.

Ward S. L., Marcus C. L., 1996 Obstructive sleep apnea in infants and young children. J Clin Neurophysiol **13**: 198–207.

Warthen D. M., Moore E. C., Kamath B. M., Morrissette J. J. D., Sanchez-Lara P. A., Sanchez P., Piccoli D. A., Krantz I. D., Spinner N. B., 2006 Jagged1 (JAG1) mutations in Alagille syndrome: increasing the mutation detection rate. Hum Mutat **27**: 436–443.

Watson A. G., Evans H. E., de Lahunta A., 1986 Ossification of the atlas-axis complex in the dog. Anat Histol Embryol **15**: 122–138.

Wayne R. K., 1986 Cranial morphology of domestic dog and wild canids: the influence of development on morphological change. Evolution **40**: 243–261.

Wayne R. K., Ostrander E. A., 2007 Lessons learned from the dog genome. Trends Genet **23**: 557–567.

Webster M., Sheets D. H., 2010 A Practical Introduction to Landmark-Based Geometric Morphometrics. In:Alroy J, Hunt G (Eds.), *Quantitative Methods in Paleobiology*, The Paleontology Society Papers, pp. 163–188.

White R. N., 2011 Surgical management of laryngeal collapse associated with brachycephalic airway obstruction syndrome in dogs. J Small Anim Pract **53**: 44–50.

Whitson S. W., 1981 Osteonectin, a bone-specific protein linking mineral to collagen. Cell **26**: 99.

Wiedmer M., Oevermann A., Borer-Germann S. E., Gorgas D., Shelton G. D., Drögemüller M., Jagannathan V., Henke D., Leeb T., 2015 A RAB3GAP1 SINE Insertion in Alaskan Huskies with Polyneuropathy, Ocular Abnormalities, and Neuronal Vacuolation (POANV) Resembling Human Warburg Micro Syndrome 1 (WARBM1). G3 (Bethesda) **6**: 255–262.

Wieland I., Jakubiczka S., Muschke P., Cohen M., Thiele H., Gerlach K. L., Adams R. H., Wieacker P., 2004 Mutations of the ephrin-B1 gene cause craniofrontonasal syndrome. Am J Hum Genet **74**: 1209–1215.

Wiese S., Reidegeld K. A., Meyer H. E., Warscheid B., 2006 Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics **7**: 340–350.

Wild A., Kalff-Suske M., Vortkamp A., Bornholdt D., König R., Grzeschik K. H., 1997 Point mutations in human GLI3 cause Greig syndrome. Hum Mol Genet **6**: 1979–1984.

Wiles B. M., Llewellyn-Zaidi A. M., Evans K. M., O'Neill D. G., Lewis T. W., 2017 Large-scale survey to estimate the prevalence of disorders for 192 Kennel Club registered breeds. Canine Genet Epidemiol **4**: 8.

Wilkie A. O., Slaney S. F., Oldridge M., Poole M. D., Ashworth G. J., Hockley A. D., Hayward R. D., David D. J., Pulleyn L. J., Rutland P., 1995 Apert syndrome results from localized mutations of FGFR2 and is allelic with Crouzon syndrome. Nat Genet **9**: 165–172.

Winograd J., Reilly M. P., Roe R., Lutz J., Laughner E., Xu X., Hu L., Asakura T., vander Kolk C., Strandberg J. D., Semenza G. L., 1997 Perinatal lethality and multiple craniofacial malformations in MSX2 transgenic mice. Hum Mol Genet **6**: 369–379.

Wolf Z. T., Leslie E. J., Arzi B., Jayashankar K., Karmi N., Jia Z., Rowland D. J., Young A., Safra N., Sliskovic S., Murray J. C., Wade C. M., Bannasch D. L., 2014 A LINE-1 insertion in DLX6 is responsible for cleft palate and mandibular abnormalities in a canine model of Pierre Robin sequence. PLoS Genetics **10**: e1004257.

Wu M., Li J., Engleka K. A., Zhou B., Lu M. M., Plotkin J. B., Epstein J. A., 2008 Persistent expression of Pax3 in the neural crest causes cleft palate and defective osteogenesis in mice. J Clin Invest **118**: 2076–2087.

Xu H., Guan J., Yi H., Yin S., 2014 A systematic review and meta-analysis of the association between serotonergic gene polymorphisms and obstructive sleep apnea syndrome. PLoS ONE **9**: e86460.

Yamahara K., Nakagawa T., Ito J., Kinoshita K., Omori K., Yamamoto N., 2017 Netrin 1 mediates protective effects exerted by insulin-like growth factor 1 on cochlear hair cells. Neuropharmacology **119**: 26–39.

Yang J., Weedon M. N., Purcell S., Lettre G., Estrada K., Willer C. J., Smith A. V., Ingelsson E., O'Connell J. R., Mangino M., Mägi R., Madden P. A., Heath A. C., Nyholt D. R., Martin N. G., Montgomery G. W., Frayling T. M., Hirschhorn J. N., McCarthy M. I., Goddard M. E., Visscher P. M., 2011 Genomic inflation factors under polygenic inheritance. Eur J Hum Genet **19**: 807–812.

Yang L., Tsang K. Y., Tang H. C., Chan D., Cheah K. S. E., 2014 Hypertrophic chondrocytes can become osteoblasts and osteocytes in endochondral bone formation. Proc Natl Acad Sci USA **111**: 12097–12102.

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**: 2865–2871.

Yisraeli J. K., 2004 VICKZ proteins: a multi-talented family of regulatory RNA-binding proteins. Biol Cell **97**: 87–96.

Yoshida C. A., Yamamoto H., Fujita T., Furuichi T., Ito K., Inoue K.-I., Yamana K., Zanma A., Takada K., Ito Y., Komori T., 2004 Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog. Genes Dev **18**: 952–963.

Yoshida T., Vivatbutsiri P., Morriss-Kay G., Saga Y., Iseki S., 2008 Cell lineage in mammalian craniofacial mesenchyme. Mech Dev **125**: 797–808.

Yousfi M., Lasmoles F., Marie P. J., 2002 TWIST inactivation reduces CBFA1/RUNX2 expression and DNA binding to the osteocalcin promoter in osteoblasts. Biochem Biophys Res Commun **297**: 641–644.

Zapata I., Serpell J. A., Alvarez C. E., 2016 Genetic mapping of canine fear and aggression. BMC Genomics **17**: 572.

Zeder M. A., 2012 Pathways to Animal Domestication. In: *Biodiversity in Agriculture Domnesitcation, Evolution and Sustainability*, pp. 227–259.

Zhang M., Pritchard M. R., Middleton F. A., Horton J. A., Damron T. A., 2008 Microarray analysis of perichondral and reserve growth plate zones identifies differential gene expressions and signal pathways. Bone **43**: 511–520.

Zhao Q., Brauer P. R., Xiao L., McGuire M. H., Yee J. A., 2002 Expression of parathyroid hormone-related peptide (PthrP) and its receptor (PTH1R) during the histogenesis of cartilage and bone in the chicken mandibular process. J Anat **201**: 137–151.

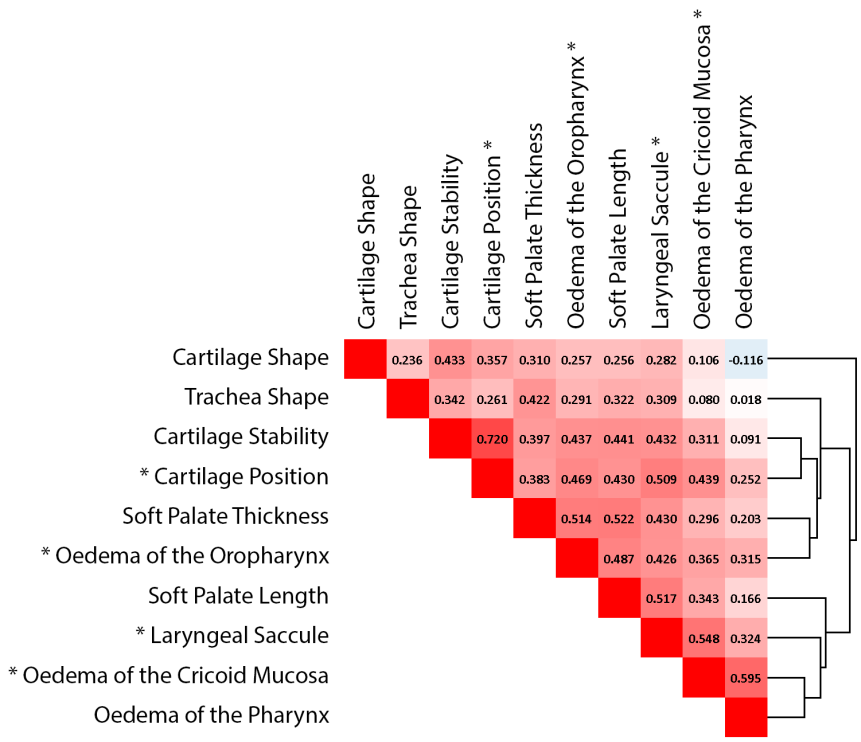
Zheng Q., Zhou G., Morello R., Chen Y., Garcia-Rojas X., Lee B., 2003 Type X collagen gene regulation by Runx2 contributes directly to its hypertrophic chondrocyte-specific expression in vivo. J Cell Biol **162**: 833–842.

Zhou X., Stephens M., 2012 Genome-wide efficient mixed-model analysis for association studies. Nat Genet **44**: 821.

Zou Q., Wang X., Liu Y., Ouyang Z., Long H., Wei S., Xin J., Zhao B., Lai S., Shen J., Ni Q., Yang H., Zhong H., Li L., Hu M., Zhang Q., Zhou Z., He J., Yan Q., Fan N., Zhao Y., Liu Z., Guo L., Huang J., Zhang G., Ying J., Lai L., Gao X., 2015 Generation of gene-target dogs using CRISPR/Cas9 system. J Mol Cell Biol **7**: 580–583.

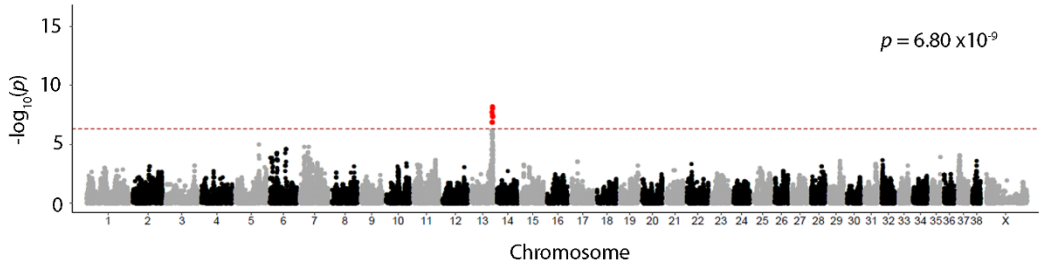
Appendix 1

Supplementary material in support of the manuscript submitted to PLoS Genetics entitled “An *ADAMTS3* Missense Variant is Associated with Norwich Terrier Upper Airway Syndrome” which has been included in Chapter 6 of this thesis.

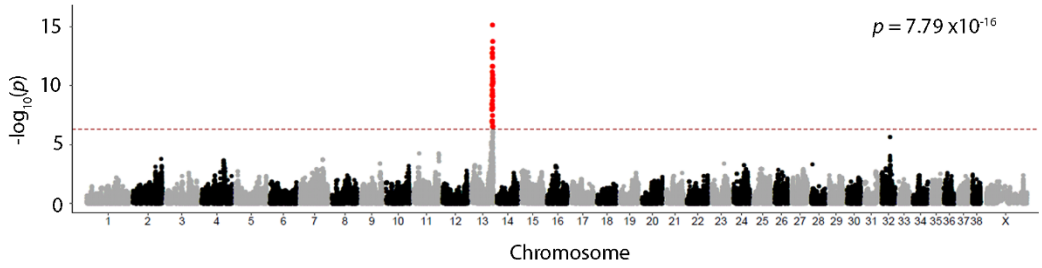


Supplemental Figure 1 - Phenotype Correlations. Pearson's correlation scores between upper airway phenotypes with dendrogram indicating relationships between phenotypes. Phenotypes returning significant associations in the GWAS are marked with (*).

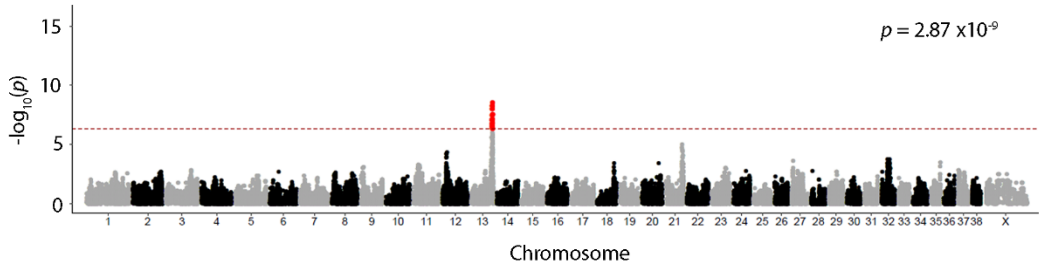
A. Cartilage Position



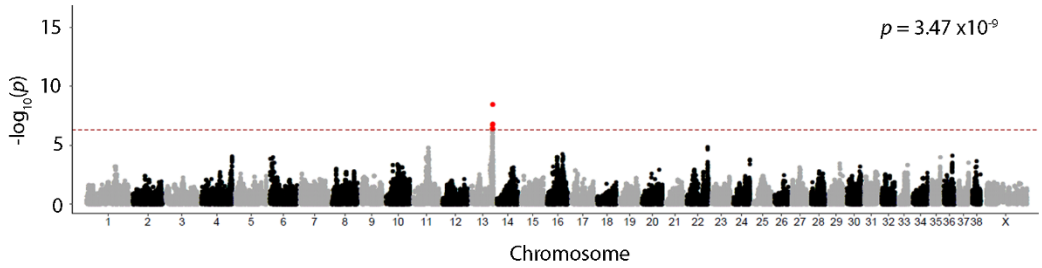
B. Laryngeal Saccule



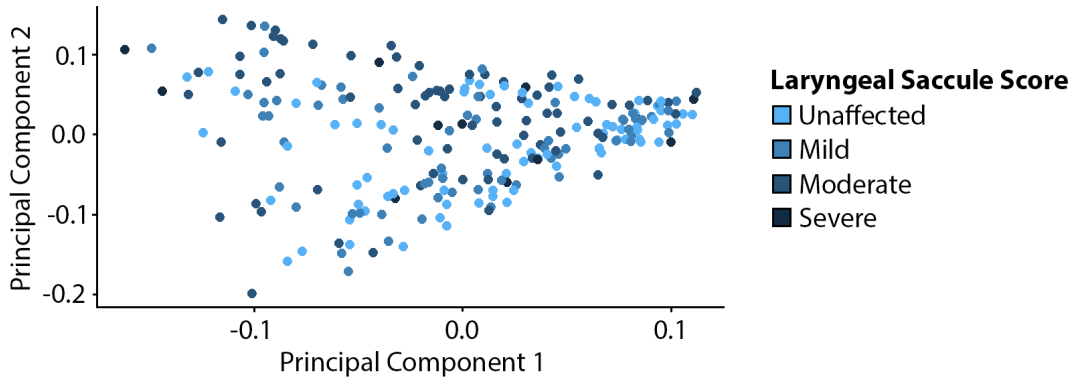
C. Oedema of the Cricoid Mucosa



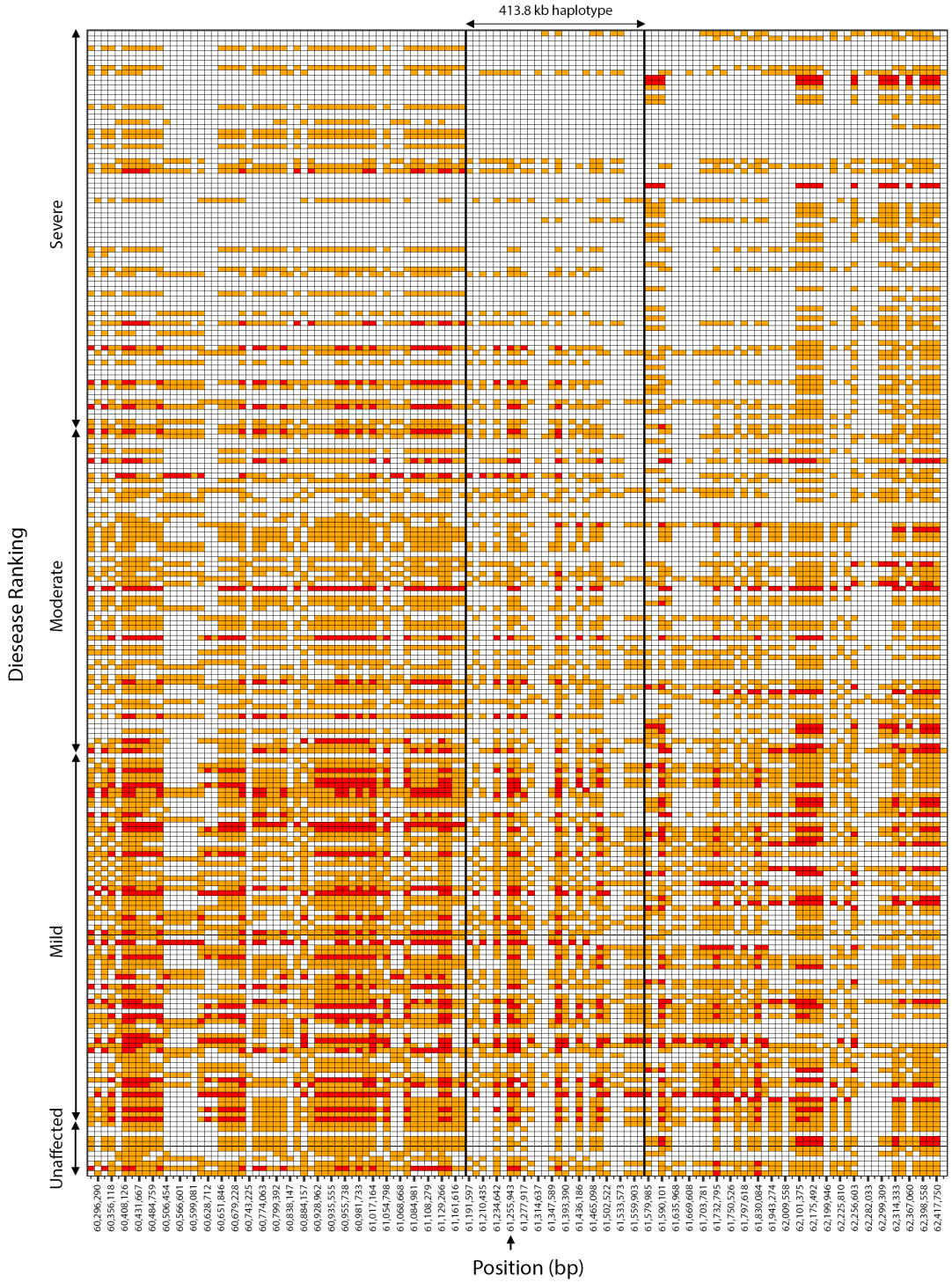
D. Oedema of the Oropharynx



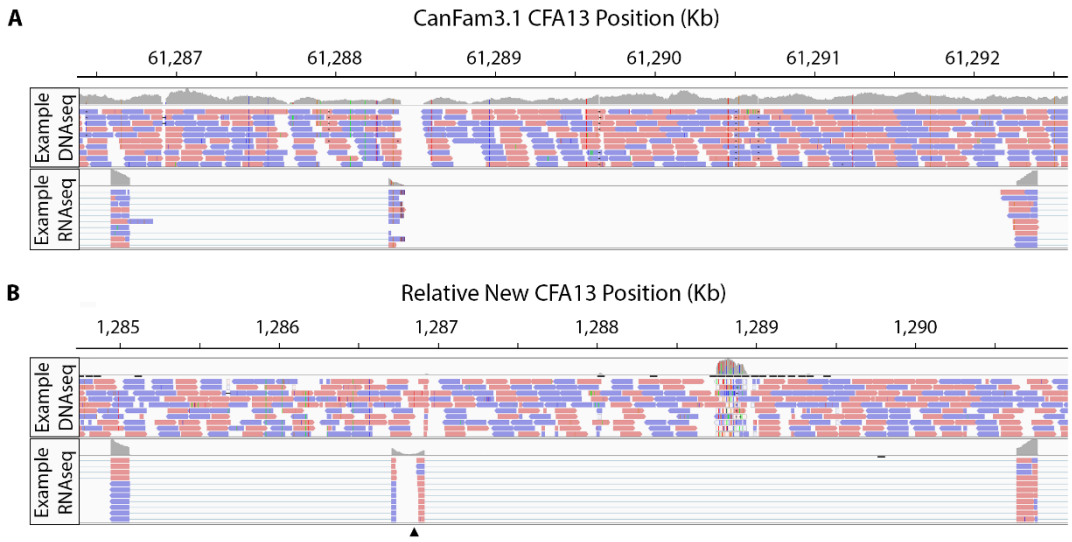
Supplemental Figure 2 - All GWAS Returning Significant Associations. Manhattan plots for (A) cartilage position, (B) laryngeal saccule, (C) oedema of the cricoid mucosa and (D) oedema of the oropharynx. The red dashed line denotes Bonferroni correction threshold (4.75×10^{-7}) and maximum significance values for index SNPs of each test are given.



Supplemental Figure 3 - Array Genotype PCA. Principal component 1 and 2 of the array genotypes do not segregate by laryngeal saccule score



Supplemental Figure 4 - Complete Haplotypes. Phased haplotypes (horizontal rows) for the region surrounding the index marker (chr13:61,255,943; arrowhead) on CFA13 for all 233 Norwich Terriers in the study cohort. Individuals are ranked by phenotype severity in order of their GWAS significance (laryngeal saccule > cartilage position > oedema of the cricoid mucosa > oedema of the oropharynx).



Supplemental Figure 5 - Error in CanFam3.1 Scaffold. DNA and RNA short-read sequencing data aligned to the (A) CanFam3.1 and (B) newly created consensus sequence for the region. Read coverage across the 5' end of exon 20 and its immediate intron is lost. The error is corrected in the new consensus as indicated by complete coverage in the region.

CFA	Position	Marker	Allele	Laryngeal Saccule		Oedema of the Cricoid Mucosa		Cartilage Position		Oedema of the Oropharynx	
				p-value	beta	p-value	beta	p-value	beta	p-value	beta
13	59,035,431	BICF2P1050096	A > G	3.43E-09	-2.5059	2.05E-07	-0.1896	2.01E-08	-0.8549		
13	59,083,075	TIGRP2P184171_rs8895564	A > C	1.10E-08	-2.4709	8.42E-08	-0.1985				
13	59,171,427	BICF2G630749007	C > G	1.10E-08	-2.4709	8.42E-08	-0.1985				
13	59,215,669	BICF2G630748960	C > T	1.10E-08	-2.4709	8.42E-08	-0.1985				
13	59,244,558	BICF2G630748919	A > G	1.10E-08	-2.4709	8.42E-08	-0.1985				
13	59,876,151	BICF2G630748126	T > C	4.10E-09	-2.4944	9.87E-09	-0.2083				
13	59,921,835	BICF2P559788	C > T					1.40E-07	-0.9191		
13	59,932,325	TIGRP2P184784_rs8637716	T > C					1.40E-07	-0.9191		
13	59,946,572	BICF2G630748106	C > G	4.10E-09	-2.4944	9.87E-09	-0.2083				
13	60,059,427	BICF2G630747989	T > C					1.40E-07	-0.9191		
13	60,104,647	BICF2S23551619	C > T	9.87E-08	-2.2869	1.62E-07	-0.1927				
13	60,119,822	BICF2P890262	C > T	9.87E-08	-2.2869	1.62E-07	-0.1927				
13	60,158,661	BICF2G630747923	A > G	9.18E-11	-2.7353	3.55E-08	-0.2009				
13	60,167,490	BICF2G630747914	A > G	9.18E-11	-2.7353	3.55E-08	-0.2009				
13	60,204,662	BICF2G630747899	C > G	6.42E-10	-2.8798	3.45E-07	-0.2050				
13	60,211,955	BICF2G630747887	A > G	1.12E-07	-2.5130						
13	60,220,792	BICF2G630747865	G > T	6.42E-10	-2.8798	3.45E-07	-0.2050				
13	60,231,218	BICF2G630747844	T > C	6.42E-10	-2.8798	3.45E-07	-0.2050				
13	60,243,398	BICF2G630747830	T > C	1.12E-07	-2.5130						
13	60,356,118	TIGRP2P184853_rs8673615	T > G	1.42E-07	-2.5290						
13	60,408,126	BICF2G630747706	C > G	1.76E-13	-3.0415	5.60E-09	-0.2087				
13	60,417,420	BICF2P940986	A > G	1.76E-13	-3.0415	5.60E-09	-0.2087				
13	60,431,667	BICF2S2376537	G > A	7.44E-12	-3.1082	2.69E-07	-0.2030				

13	60,474,403	TIGRP2P184870_rs9068575	C > T	2.38E-10	-2.9146					4.18E-07	-0.3842
13	60,484,759	BICF2G630747649	C > T	7.89E-09	-2.7127						
13	60,496,347	BICF2G630747628	C > T	7.68E-11	-2.7362						
13	60,651,846	BICF2G630747522	A > G	7.89E-09	-2.7127						
13	60,656,371	BICF2G630747517	A > G	7.89E-09	-2.7127						
13	60,679,228	BICF2G630747483	C > T	7.89E-09	-2.7127						
13	60,707,730	BICF2G630747446	T > C	2.41E-11	-2.7528	4.26E-07	-0.1806				
13	60,884,157	TIGRP2P184965_rs8996044	G > T	1.98E-09	-3.0824			9.79E-09	-1.0521		
13	60,928,962	BICF2G630747127	T > C	8.14E-10	-2.8460						
13	60,934,509	BICF2G630747111	T > C	8.14E-10	-2.8460						
13	60,935,555	BICF2S23032685	A > G	8.14E-10	-2.8460						
13	60,948,625	BICF2G630747091	T > C	2.59E-12	-2.8867	4.11E-07	-0.1825				
13	60,955,738	BICF2G630747083	A > G	2.59E-12	-2.8867	4.11E-07	-0.1825				
13	60,974,872	BICF2G630747065	G > A	8.14E-10	-2.8460						
13	60,981,733	BICF2S23549893	G > A	2.59E-12	-2.8867	4.11E-07	-0.1825				
13	60,990,552	BICF2G630747041	A > C	4.92E-10	-2.8464					1.98E-07	-0.3919
13	61,017,164	BICF2G630747018	G > A	1.92E-13	-2.9879	7.70E-08	-0.1900				
13	61,039,571	BICF2G630746980	T > A	3.49E-08	-2.9012			3.37E-08	-1.0417		
13	61,054,798	TIGRP2P184991_rs9060216	T > C	4.21E-10	-2.8607						
13	61,076,409	BICF2G630746923	G > A	2.72E-10	-2.9433						
13	61,129,266	TIGRP2P185016_rs8744939	G > T	6.84E-14	-3.0560	2.77E-08	-0.1976				
13	61,139,969	BICF2P1046807	C > T	4.56E-13	-2.9556	1.59E-07	-0.1862				
13	61,234,642	TIGRP2P185072_rs9078356	T > C	7.80E-11	-2.8691			4.17E-08	-0.8752		
13	61,255,943	TIGRP2P185081_rs9043975	G > A	7.79E-16	-3.2283	2.87E-09	-0.2076	8.79E-09	-0.8536		
13	61,270,000	TIGRP2P185090_rs8834045	A > C	3.76E-13	-3.2052			6.80E-09	-0.9351		
13	61,324,727	BICF2G630746664	C > A	1.30E-11	3.7540					3.47E-09	0.5585

13	61,390,999	BICF2G630746542	C > T	1.92E-14	-3.1259	2.93E-08	-0.1973	4.62E-08	-0.8236		
13	61,418,089	BICF2G630746493	T > G	5.21E-11	-2.9697						
13	61,436,186	BICF2G630746463	T > G	4.52E-11	-3.0291						
13	61,496,628	BICF2P555379	T > C	2.94E-11	-2.9468						
13	61,549,154	BICF2S2355385	G > T	2.92E-07	-3.3241						
13	61,567,707	BICF2G630746277	T > C	2.92E-07	-3.3241						
13	61,732,795	BICF2G630746093	A > G	4.98E-11	-3.0130					4.34E-07	-0.3826
13	61,830,084	BICF2G630745919	A > G	6.71E-11	-3.0095					1.64E-07	-0.3965

Supplementary Table 1 - Significant SNVs. Four phenotypes have markers that surpass a Bonferroni correction threshold which correspond to the same QTL on CFA13. Only *p*-values exceeding this threshold (4.76×10^{-7}) are reported. Alleles are reported as ancestral > derived.

Position (bp)	Reference	Alternate
61170181	A	G
61189929	A	G
61190053	A	G
61190184	T	C
61190246	C	G
61190348	AATG	TATT
61190752	CCTCT	CCT,C
61190789	GGTGAATAAAC	CGTGAATAAAT,TGTGAATAAAT
61191362	T	A
61191408	GAGA	G
61191554	T	C
61191750	A	G
61192115	C	T
61192271	A	T
61192476	G	A
61192504	G	A
61192540	G	T
61193292	T	C
61193340	G	C
61193366	G	T
61193672	C	A
61193693	G	T
61193761	C	T
61194302	C	T
61194342	C	T
61194542	G	A
61194604	G	A
61194613	CCTTTTTT	C
61195300	A	G
61195437	C	A
61195466	A	G
61195503	C	T
61195570	A	T
61195787	TTTTTTTTTTTTA	TATTTTTTTTA
61195847	T	TTT
61195902	TCAAAACCTA	CCAAAACCTT
61195965	G	A
61196420	A	C
61196508	AG	GA,GG
61197002	C	T
61197349	C	T

61197483	A	G
61197836	T	C
61197933	A	AT
61197997	TAACTG	CAACTA
61198419	G	A
61205561	G	GT,GTT
61207979	G	T
61208736	G	A
61210349	G	T
61212738	T	A
61212777	GGT	G,GGTGTGT,GGTGTGTGT
61213741	C	T
61213959	G	A
61214054	AAT	A
61215142	G	C
61216135	CG	C
61216488	A	T
61218653	G	A
61219883	TTTTG	T
61220575	C	G
61220670	A	G
61221130	C	G
61221146	C	T
61221333	AAATAAATG	A
61221608	A	G
61221652	AGAAAATAATATTAG	CGAAAATAATATTAA
61222124	G	C
61222427	TGTATAACGTT	TATAACGC
61224152	A	G
61226001	G	A
61227517	C	T
61227956	A	C
61228172	C	T
61228229	T	C
61228323	G	A
61228409	T	C
61228496	TATAG	T
61228702	AAATAAAAAAATAAAAAAAT	TAAAAAAAAATTAAAAAAAA
61228935	G	GTT,GTTT
61229680	G	A
61230145	G	T
61230167	A	G

61230191	T	TA
61230651	C	A
61231263	G	C
61241538	T	C
61243071	T	TA
61244812	C	T
61244832	T	A,C
61244946	G	C
61245474	T	C
61246597	G	A
61247277	T	A
61247324	G	C
61248328	T	G
61248415	C	CT
61250069	AT	A
61250267	A	C
61250391	C	T
61251419	TGGAGGGAGGGAGAGAGGGAG	T,TGGAGGGAGGGAGGGAGGGAGAGAGGGAG
61252577	C	T
61252689	TCTG	T
61252910	A	AG
61254081	CATGT	AATGG
61255943	A	G
61256419	T	A
61258164	AGTGGCCTAA	GGTGGCCTAA,GGTGGCCTAC
61258363	T	C
61258744	C	T
61258774	CTT	C
61258868	T	G
61258906	ATGA	TCTT
61259076	T	G
61259876	C	A
61262188	T	C
61262517	AG	CA,CG
61262834	G	C,T
61263699	A	T
61263848	A	G
61264085	TAAGTTAGAAGATAG	CAAGTTAGAAGATAA
61264492	C	G
61265325	G	T
61265383	AAG	GAA
61267404	C	T

61271979	T	C
61272088	T	G
61272563	CACAA	TACAA,TACAT
61272600	T	C
61274339	C	T
61274752	AATGGGC	GATGGGC,GATGGGT
61274854	G	GA
61274950	C	T
61276616	ACT	A
61276989	TGACAAATAAAC	AGACAAATAAAA,AGACAAATAAAC
61278501	C	T
61281451	C	G
61281492	G	A
61281530	C	T
61281624	CATTAGAAACAACAAA TACCCAC	TATTAGAAATGACAAATACCCGA
61281694	A	G
61282730	CATGGCAAGA	TATGGCAAGG
61282879	T	TAAA
61283039	T	TTTTTATTTTATTTATTTTATTTTA
61283136	T	TTTTTTGTTG,TTTTTTTGTTGTTG
61283508	GGAAAAA	GAG
61284432	C	A
61285186	T	C
61285305	T	TAA,TA
61286095	AATTTTT	A
61286424	A	G
61286641	G	T
61287796	G	A
61287813	A	T
61290789	T	C
61306616	ATT	A,AT
61331648	T	A
61373554	CGTTT	CGTTTGTTT,C
61376126	CGCAT	C
61376261	CACAAATATTGT	GACAAATATTGC,GACAAGTATTGT
61376333	C	CAGAG
61377534	CTGGTTGTG	TTGGTTGTA
61377776	T	C
61378126	G	A
61378382	C	T
61381005	TTCTC	TCTCTC,T
61381205	T	A

61381670	T	A
61382048	T	TAA
61384776	TGTGGGGGTGAATGGGTGACG	GGTGGGGGTGACTGGGTGACC,GGTGGGGGTGACTGGGTGACG
61385160	T	TTTTA
61392816	CGTTTAGTGGTTTGGCG	GGCTTAGTGGTTTGGCA,GGCTTAGTGGTTTGGCG,GGCTTAGTGGTTTGGTG
61393645	T	C
61394050	G	T
61394062	G	A
61394068	GAAGA	G
61394075	C	A
61406265	CTT	C
61407704	TGTTCAGAA	CGTTCAGAA,CGTTCAGAC
61407920	A	G
61407998	G	T
61408820	G	A
61411415	C	T
61412572	G	GA
61415543	T	C
61415634	TTTTTTTTTTT	GTTTTGTTTTG
61415991	G	A
61416503	G	GT
61419324	TTTC	T
61420218	C	CAATAAATAAATA,CAATAAATAAATAAATA
61420225	CAAAC	TAAAC,TAAAT
61421201	TA	TAA,T
61422615	T	C
61428232	G	A
61428302	G	A
61428495	G	A
61429847	A	AAGAGAGAGAG,AAGAGAGAGAGAGAGAG
61431382	GCCTGCAGCCA	ACCTGCAGCCG
61433428	TGATTCAAT	CGATTCAAC
61434837	CA	C
61437360	C	A
61438274	C	T
61439951	G	GA
61440216	C	CAT
61440933	C	A
61441283	AG	CT
61441819	T	C
61442856	AT	A

61443055	TAA	TAAA,T
61443079	A	AT
61443476	C	T
61444195	CTGACCCTTT	TTGACCCTTG
61445225	C	G
61445273	T	A
61445408	C	T
61445463	A	G
61446470	G	A
61448221	G	A
61448429	G	A
61448704	G	A
61449347	C	T
61449742	C	T
61449822	A	G
61449918	TGCCATGAGGACCG	CGCCATGAGGACCA
61460566	G	C
61461747	GGATAGATA	G,GGATAGATAGATA
61462213	TTTTTTTTTT	ATTTTTTTTA
61462620	T	C
61462680	GATCCCCAGGTTAAGATC	AATTCCCAGGTTAAGATT
61463119	G	A
61492288	TCTAGTACATTT	ACTAGTACATTC,ACTAGTACATTT

Supplementary Table 2 - List of Remaining Variants. Variants passing filtering criteria when aligned to the CanFam3.1 and new consensus (Zoey2.3) assemblies.

Breed	n	A/A	G/A	G/G	N/N	Disease Allele Frequency
Norwich Terrier	401	132 (32.9)	195 (48.6)	72 (18)	2 (0.5)	0.572
Bulldog	24	19 (79.2)	3 (12.5)	2 (8.3)	0 (0)	0.854
French Bulldog	99	3 (3)	17 (17.2)	79 (79.8)	0 (0)	0.116
Staffordshire Bull Terrier	8	0 (0)	2 (25)	6 (75)	0 (0)	0.125
German Spitz (Mittel)	8	0 (0)	1 (12.5)	7 (87.5)	0 (0)	0.063
Pomeranian	8	0 (0)	1 (12.5)	7 (87.5)	0 (0)	0.063
Afghan Hound	1	0 (0)	0 (0)	1 (100)	0 (0)	0
Airedale Terrier	1	0 (0)	0 (0)	1 (100)	0 (0)	0
Akita	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Alaskan Malamute	1	0 (0)	0 (0)	1 (100)	0 (0)	0
Alpine Dachsbracke	8	0 (0)	0 (0)	8 (100)	0 (0)	0
American Bulldog	2	0 (0)	0 (0)	2 (100)	0 (0)	0
Appenzeller Cattle Dog	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Australian Cattle Dog	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Australian Kelpie	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Australian Shepherd	18	0 (0)	0 (0)	18 (100)	0 (0)	0
Barbet	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Basset Hound	7	0 (0)	0 (0)	7 (100)	0 (0)	0
Bavarian Mountain Hound	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Beagle	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Bearded Collie	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Bedlington Terrier	4	0 (0)	0 (0)	4 (100)	0 (0)	0
Belgian Shepherd (Miniature Spitz)	7	0 (0)	0 (0)	7 (100)	0 (0)	0
Bernese Mountain Dog	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Bichon Frise	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Border Collie	17	0 (0)	0 (0)	17 (100)	0 (0)	0
Borzoi	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Boston Terrier	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Boxer	8	0 (0)	0 (0)	7 (87.5)	1 (12.5)	0
Bull Terrier	7	0 (0)	0 (0)	7 (100)	0 (0)	0
Bullmastiff	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Cairn Terrier	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Cão da Serra de Aires	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Cavalier King Charles Spaniel	10	0 (0)	0 (0)	10 (100)	0 (0)	0
Chesapeake Bay Retriever	1	0 (0)	0 (0)	1 (100)	0 (0)	0
Chinese Crested Dog	8	0 (0)	0 (0)	8 (100)	0 (0)	0

Chinese Shar-Pei	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Chow Chow	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Collie	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Curly Coated Retriever	7	0 (0)	0 (0)	7 (100)	0 (0)	0
Czechoslovakian Wolfdog	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Dachshund	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Dalmatian	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Doberman	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Dogue de Bordeaux	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Elo	8	0 (0)	0 (0)	8 (100)	0 (0)	0
English Cocker Spaniel	9	0 (0)	0 (0)	9 (100)	0 (0)	0
English Pointer	8	0 (0)	0 (0)	8 (100)	0 (0)	0
English Springer Spaniel	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Entlebucher Cattle Dog	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Eurasier	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Flat Coated Retriever	9	0 (0)	0 (0)	9 (100)	0 (0)	0
German Hound	1	0 (0)	0 (0)	1 (100)	0 (0)	0
German Hunting Terrier	8	0 (0)	0 (0)	8 (100)	0 (0)	0
German Shepherd Dog	8	0 (0)	0 (0)	8 (100)	0 (0)	0
German Short-haired Pointer	5	0 (0)	0 (0)	5 (100)	0 (0)	0
German Spaniel	9	0 (0)	0 (0)	9 (100)	0 (0)	0
German Spitz (Giant Spitz)	6	0 (0)	0 (0)	6 (100)	0 (0)	0
German Spitz (Keeshound)	8	0 (0)	0 (0)	8 (100)	0 (0)	0
German Spitz Klein	7	0 (0)	0 (0)	7 (100)	0 (0)	0
German Wirehaired Pointer	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Giant Schnauzer	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Golden Retriever	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Gordon Setter	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Great Dane	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Greater Swiss Mountain Dog	1	0 (0)	0 (0)	1 (100)	0 (0)	0
Greyhound	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Hovawart	5	0 (0)	0 (0)	5 (100)	0 (0)	0
Irish Red Setter	1	0 (0)	0 (0)	1 (100)	0 (0)	0
Irish Setter	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Irish Terrier	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Jack Russel Terrier	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Kromfohrländer	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Labrador Retriever	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Lagotto Romagnolo	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Lakeland Terrier	6	0 (0)	0 (0)	6 (100)	0 (0)	0
Landseer	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Leonberger	9	0 (0)	0 (0)	9 (100)	0 (0)	0

Lhasa Apso	14	0 (0)	0 (0)	13 (92.9)	1 (7.1)	0
Maltese	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Mastiff	1	0 (0)	0 (0)	1 (100)	0 (0)	0
Miniature Bull Terrier	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Miniature Pincscher	7	0 (0)	0 (0)	7 (100)	0 (0)	0
Miniature Schnauzer	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Mixed	2	0 (0)	0 (0)	2 (100)	0 (0)	0
Mudi	7	0 (0)	0 (0)	7 (100)	0 (0)	0
Newfoundland	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Norfolk Terrier	12	0 (0)	0 (0)	17 (100)	0 (0)	0
Nova Scotia Duck Tolling Retriever	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Papillon	1	0 (0)	0 (0)	1 (100)	0 (0)	0
Pekingese	2	0 (0)	0 (0)	2 (100)	0 (0)	0
Perro de Agua Espanõl	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Peruanischer Hairless Dog	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Picardy Sheepdog	1	0 (0)	0 (0)	1 (100)	0 (0)	0
Polish Lowland Sheepdog	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Poodle	14	0 (0)	0 (0)	14 (100)	0 (0)	0
Pug	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Rhodesian Ridgeback	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Rottweiler	9	0 (0)	0 (0)	9 (100)	0 (0)	0
Saluki	1	0 (0)	0 (0)	1 (100)	0 (0)	0
Schapendoes	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Scottish Terrier	17	0 (0)	0 (0)	17 (100)	0 (0)	0
Shetland Sheepdog	7	0 (0)	0 (0)	7 (100)	0 (0)	0
Shih Tzu	13	0 (0)	0 (0)	13 (100)	0 (0)	0
Siberian Husky	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Sloughi	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Slovak Rough-Haired Pointer	1	0 (0)	0 (0)	1 (100)	0 (0)	0
St. Bernard	7	0 (0)	0 (0)	7 (100)	0 (0)	0
Standard Schnauzer	1	0 (0)	0 (0)	1 (100)	0 (0)	0
Tibetan Terrier	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Vizsla	9	0 (0)	0 (0)	9 (100)	0 (0)	0
West Highland White Terrier	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Whippet	9	0 (0)	0 (0)	9 (100)	0 (0)	0
White Swiss Shepherd Dog	8	0 (0)	0 (0)	8 (100)	0 (0)	0
Yorkshire Terrier	1	0 (0)	0 (0)	1 (100)	0 (0)	0

Supplementary Table 3 - Multiple Breeds Genotyped for *ADAMTS3* Variant. ADAMTS3:c.2786G>A genotypes of 1,353 dogs from up to 114 different breeds. The total number of dogs genotyped per breed (n) is given and the frequency of genotypes within breeds is given in brackets as a percentage.