The Genetics of Canine Atopic Dermatitis

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Shona Hiedi Wood.

May 2010

Acknowledgements

I would like to thank all my supervisors; Stuart Carter, Bill Ollier, Tim Nuttall and Neil McEwan, for their support and help throughout my PhD. My thanks also go to all the staff at the Centre for Integrated Genomic Medical Research, University of Manchester and the Veterinary Sciences Faculty, University of Liverpool.

Special thanks to Dylan Clements for his technical, statistical and moral support during in my PhD. I also extend my gratitude to Annette Barnes who supported me in my PhD application.

I would like to thank my fiancé Paul Thompson for standing by me throughout my PhD. He has put my career first and given me unwavering support. I hope he realises how much I appreciate and love him.

I would like to thank my parents; Donna and Mark Wood, without their love and encouragement of my curiosity I would not be where I am today. Thank you Mum and Dad for always letting me explore.

To thank my Mum, Dad and Paul I am dedicating this thesis to them.

Contents

List of figures and Tables

Abstract: The Genetics of Canine Atopic Dermatitis

Shona Hiedi Wood

Canine atopic dermatitis (cAD) is a common and severe pruritic, inflammatory skin disease that can be considered a naturally-occurring, spontaneous model of human Atopic Dermatitis (hAD). The genetics of cAD are poorly understood and therefore the aim of this project was to investigate the genetic factors involved in the pathogenesis of cAD and to identify specific gene associations with cAD within and between dog breeds. It was hoped that this study would further strengthen the evidence that dogs are a suitable model hAD.

Using dogs as a model to study the genetic basis of AD is advantageous because dog breeds form genetically isolated populations exhibiting strong linkage disequilibrium (LD). In contrast to humans where LD across the genome is weak (10-100 kb), domestic dog breeds have strong LD which extends over long distances (0.8 -5 Mb). This is highly advantageous in genetic studies because fewer genetic markers and smaller sample sizes are needed to find disease associations in dogs.

To study the genetic basis of cAD a dual approach of candidate gene association study and genome wide association study (GWAS) was used. This gave not only a novel unbiased approach but also used information from previous studies on which gene selection was based. Therefore increasing the likelihood that the causative genes involved in cAD pathogenesis could be identified.

This thesis demonstrated altered mRNA expression in 54 genes out of 22,000 transcripts by mRNA microarray in cAD. Further to this qPCR was used to confirm the microarray results and quantify gene expression in potential cAD candidate genes. This approach identified 11 genes with altered expression in cAD. The qPCR results were further correlated

with the clinical outcomes: Canine Atopic Dermatitis and Severity Index (CADESI-03) and number of responses on intra-dermal allergen tests.

Eleven novel SNPs and 1 novel microsatellite were identified by transgenomic WAVE analysis. The microsatellite was further typed in 659 dogs and but no association with cAD was found.

A GWAS with 22,362 SNPs was performed. The significant results were validated by Sequenom along with the SNPs from the candidate gene study (literature selected genes) in 659 dogs across 8 breeds. In total, 232 SNPs across 54 genes and 41 intergenic regions were genotyped on Sequenom. From this 45 putative associations were found in various breeds. A large number of these associations had relevant functions to AD and/or previous association with hAD.

Chapter 1

Literature Review

1. Literature Review

1.1 Introduction

Canine atopic dermatitis (cAD) is a common and severe puritic, inflammatory skin disease that can be considered a naturally-occurring, spontaneous model of atopic dermatitis/eczema in humans (hAD). In both dogs and humans AD can be differentiated into the more common allergen-mediated form of the disease (termed extrinsic AD in humans) and a less common non-allergen mediated form (termed intrinsic AD in humans and atopic-like dermatitis [ALD] in dogs)[1;2].

There are many animal models which display clinical signs consistent with hAD but dogs are of particular interest as they share the same environment as humans, unlike the commonly used rodent models. The pathogenesis of the disease in both humans and dogs is strongly associated with immunological hyper-reactivity, although skin barrier function, microbial colonisation and infection also have a role[3;4]. The prevalence of hAD over the last three decades has increased two to three fold in industrialised countries[5], it has been suggested that this could also be the case in cAD, however studies assessing cAD prevalence have been contradictory and biased[6].

1.2 Clinical descriptions

CAD exhibits characteristic clinical features. The typical age of onset is 6 months to 3 years[7], though it is possible that signs will present in dogs less than 6 months and greater than 7 years old[8;9]. Reports of sex bias are variable and contradictory[8;10]. Breed pre-dispositions have also been observed in cAD (detailed in section 1.6). Clinical signs of cAD can be both

seasonal and non-seasonal depending on which allergens are involved, but ALD is usually constant throughout the year[7].

Pruritus (itching) is a hallmark of cAD. This can occur in any one or all of these locations: around the face, ears, paws, flexor surfaces and/or the ventrum. Generalized pruritus is also common, although it has been noted that involvement of the dorsal lumbar skin and pinnal margins is characteristic of flea allergic dermatitis and *Sarcoptes* infestation respectively[11].

The primary lesions of cAD mainly consist of diffuse erythema of affected skin, although macular and papular erythema can be seen. Some dogs do not present with primary lesions at all and it has been suggested that primary lesions are only present when there is secondary infection[7].

Secondary lesions are also common at pruritic sites in cAD; these are non-specific and are thought to reflect chronic pruritus, inflammation and trauma. These include red-brown saliva staining, excoriations, self induced alopecia, hyper-pigmentation, scaling and lichenification.

Atopic otitis externa and conjunctivitis are both common in cAD. Moreover atopic skin is readily colonized with staphylococci and *Malassezia* compared to healthy skin[12], and therefore secondary skin and ear infections frequently complicate cAD. Clinical signs include surface overgrowth, superficial folliculitis, otitis, pododermatitis, seborrhea and hyperhidrosis[7].

Management of cAD is difficult as the pathogenesis is complex and evolving[13]. Allergen specific immunotherapy (ASIT) is recommended to induce tolerance and prevent recurrence[14]. Other topical and systemic treatments include emollient shampoos, glucocorticoids[15], antihistamines[16], ciclosporin[17] and essential fatty acids[18]. However, many

of these treatments can be time consuming, expensive, have adverse side effects and are not always effective in all patients[13].

1.3 Pathology

Much of the pathology associated with allergic diseases such as AD is a consequence of long term chronic allergic inflammation[19;20]. Lesional skin samples from cAD often exhibit varying degrees of epidermal hyperplasia, orthokeratotic or parakeratotic hyperkeratosis, hypergranulosis, spongiosis, melanosis and leukocyte exocytosis[21]. The degree of epidermal spongosis, acanthosis and hyperkeratosis is related to the severity of the clinical signs[22]. A superficial to deep, peri-vascular to diffuse dermal infiltration of CD4+ and CD8+ T-cells, monocytes, eosinophils, neutrophils, mast cells and plasma cells, which correlates with clinical severity is often seen[22]. Epidermal Langerhans cells are common and tend to group as micro-aggregates in the epidermis. The epidermis exhibits increased numbers of IgE positive Langerhans cells. Dermal mast cell hyperplasia has also been observed, and, moreover, the dermal mast cells are also frequently IgE positive[22]. γδ-T-cells, which are rare in humans, are seen in the epidermis and dermis in cAD. Whether they have any activity as effector or regulatory cells is unknown[23]. Eosinphils are also present in the dermis in cAD, and degranulated cells are also present and may be overlooked[22].

As in hAD, non-lesional skin in cAD is characterised by superficial peri-vascular infiltrates of lymphocytes, monocytes, dendritic cells and mast cells[23;24], with mild epidermal spongiosis sometimes present[22;23].

1.4 Atopic dermatitis in humans and dogs: A biphasic model?

AD may be considered an unusual disease as it has immunological abnormalities that do not fit the models of other classical immunological diseases[25]. In AD increased levels of allergen specific IgE and epidermal eosinophils are often observed; this immunological profile would suggest that AD is driven by an allergic-type Th2 response. However, it is also reported that INF-γ levels are increased in AD suggesting a Th1 response[26;27]. This observation led to the theory of sequential activation or the biphasic model of AD pathogenesis[26]. Based on this model, it is suggested that AD is an abnormal, Th2 driven over-reaction to allergens that proceeds into a chronic phase mediated by Th1 responses. It is thought that percutaneously absorbed allergens are taken up and processed by epidermal Langerhans cells[28]. These cells then present the processed allergens via MHC class II molecules to naïve T-cells[29]. There is evidence from both canine and human AD patients that this results in the activation of allergen specific T-helper2 (Th2)-like CD4+ T cells and the secretion of a Th2-type cytokine profile, which ultimately results in the production of allergen-specific IgE[30;31]. Allergen specific IgE readily binds to the high affinity receptor, FCεRI, expressed on cutaneous Langerhans and mast cells[32]. Binding of the allergen triggers mast cell degranulation and the release of inflammatory mediators such as leukotrienes, prostaglandins, proteases, various cytokines and chemokines, and histamine resulting in the classic acute inflammatory response seen in allergy. However, infiltrating eosinophils release IL-12, which leads to a Th1 mediated cell-mediated immune response and formation of chronic AD lesions[7;33]. This process is summarised in Figure 1.1.

Figure 1.1 Diagram of Th Cell and Cytokine Responses in Atopic Dermatitis Based on the Biphasic model

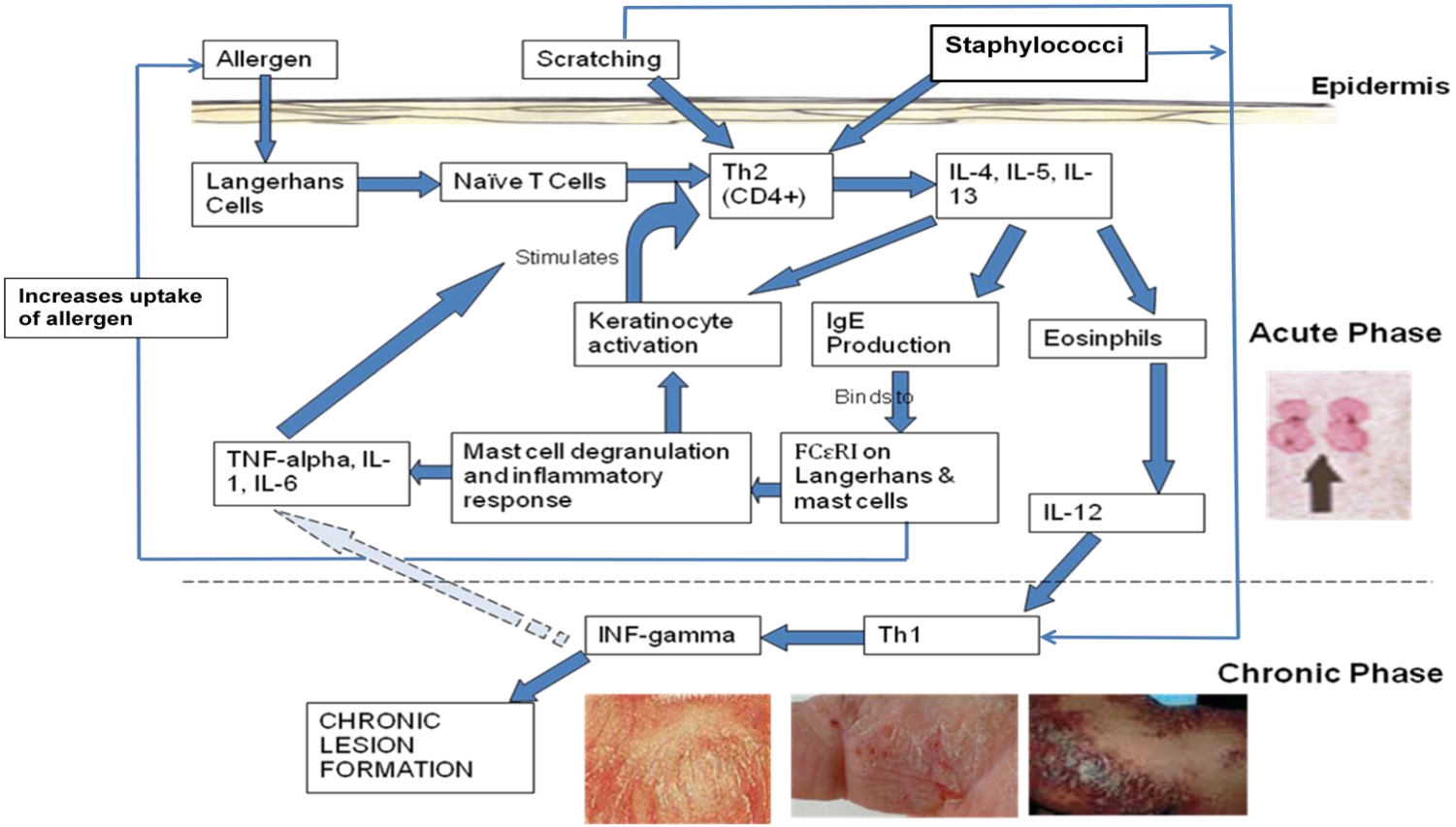


Figure 1.1: The allergens are absorbed through the skin and taken up by Langerhans cells, and presented via MHC class II to naïve T cells these become Th2 cells. These Th2 cells release classic allergy cytokines. These cytokines cause keratinocyte activation, IgE production and eosinophil infiltration into the epidermis. Keratinocytes stimulate the release of more Th2 cells compounding the response. IgE binds to receptors on Langerhans and mast cells; this causes degranulation and an inflammatory response that further activates keratinocytes to stimulate the release of Th2. Subsequently TNF-alpha, IL-1 and IL-6 further compound the Th2 response. This represents the acute phase of AD with little or no lesions but a lot of inflammation and itching are seen. The eosinophils cause the release of IL-12 which leads to a Th1 response. Th1 stimulates INF-gamma release; this is thought to lead to chronic lesion formation. Concurrently TNF-alpha is still producing a Th2 response, which is further compounded by scratching and microbial colonisation with Staphylococci.

Intrinsic human AD and canine ALD are, however, not associated with allergen specific IgE. These conditions are nevertheless clinically similar to extrinsic AD in humans and 'classical' AD in dogs[1;9;34-36]. It is therefore likely that the pathogenesis is more complex than a simple Th2/Th1 switch, and that other factors contribute to the pathogenesis.

1.5 Atopic Dermatitis as a complex disease

In humans and dogs, over the last few years, it has become well established that AD is a complex disease with many factors contributing to the pathogenesis[37;38]. In the following sections I will discuss how skin barrier function, innate and adaptive immunity, microbial colonization and infection are all thought to contribute to the biphasic response and therefore the pathology of AD.

Mechanical barrier: Epidermal dysfunction

Allergic inflammation in AD is thought to be associated with sensitization to normally innocuous environmental allergens. Sensitization facilitates the Th2 response seen in Figure 1.1. Development of clinically significant sensitization can be affected by: individual susceptibility, allergen type/concentration, presence of endotoxin and/or chitin, and frequency and/or route of allergen challenge[39]. The route of allergen challenge is of interest in AD as the permeability of the epidermal barrier can be affected by genetic and environmental factors that increase allergen penetration and potential sensitization. Moreover, it has been demonstrated that allergen challenge by this route favours a Th2 response[37;40;41]. Consequently it has been proposed that AD is the result of faulty epidermal barrier protection [42;43] (see Figure 1.2), allowing either more allergens to pass the epidermal barrier and/or increasing sensitization to these allergens. A potential mechanism for the development of non-IgE associated disease (intrinsic AD/ALD) involving the skin barrier has also been put forward (Figure 1.2)[44].

Figure 1.2 Disruption of the Skin Barrier and Different Routes of Allergen Presentation (adapted from Callard *et al* [44])

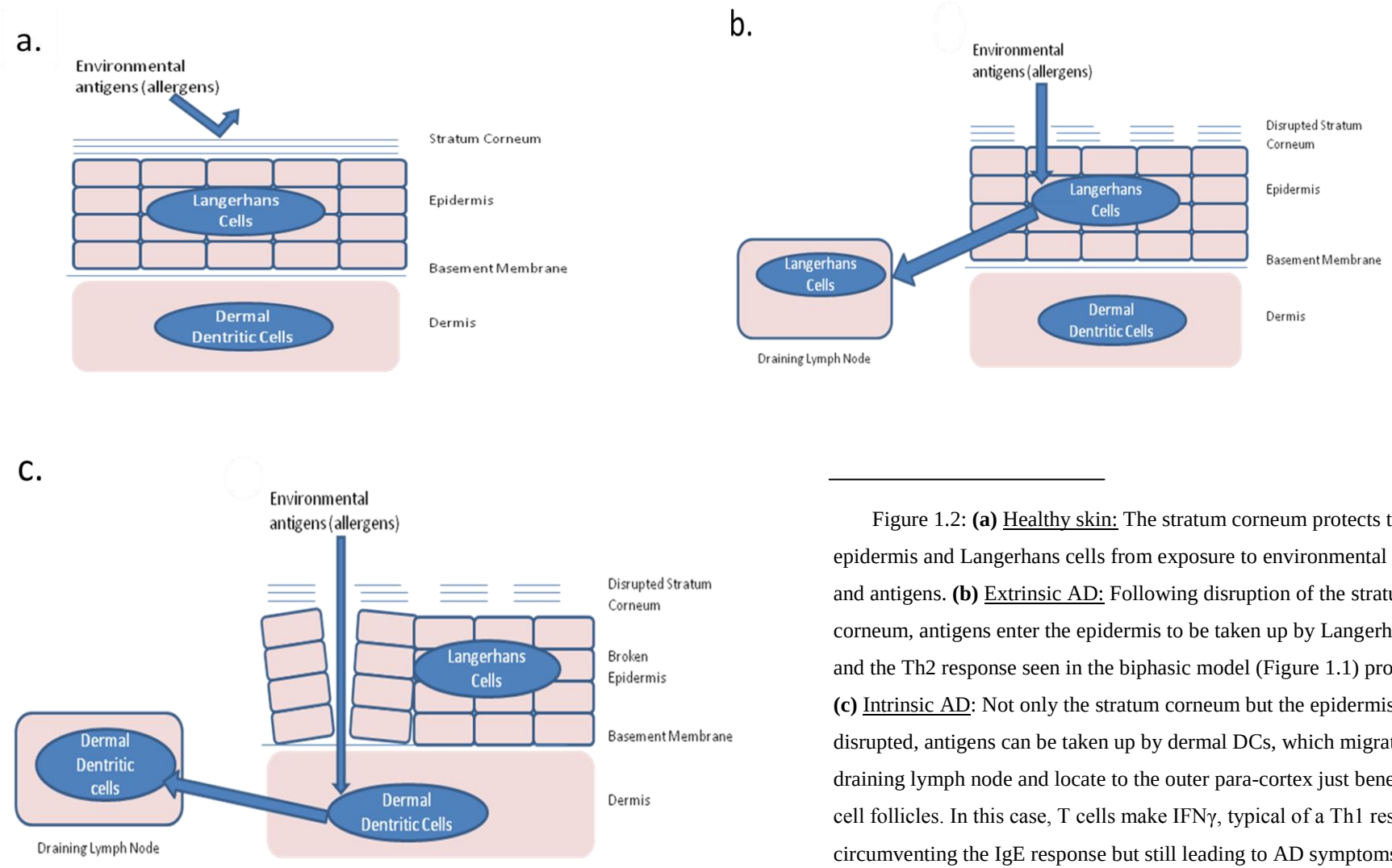


Figure 1.2: **(a)** Healthy skin: The stratum corneum protects the epidermis and Langerhans cells from exposure to environmental chemicals and antigens. **(b)** Extrinsic AD: Following disruption of the stratum corneum, antigens enter the epidermis to be taken up by Langerhans cells and the Th2 response seen in the biphasic model (Figure 1.1) proceeds. **(c)** Intrinsic AD: Not only the stratum corneum but the epidermis is disrupted, antigens can be taken up by dermal DCs, which migrate to the draining lymph node and locate to the outer para-cortex just beneath the B cell follicles. In this case, T cells make IFNγ, typical of a Th1 response, circumventing the IgE response but still leading to AD symptoms.

Evidence to support the theory of epidermal barrier dysfunction in humans is extensive. An earlier study demonstrated that trans-epidermal water loss (TEWL) at lesional sites on human atopic patients was twice that at non-lesional sites and four times that of healthy skin[45]. In the same study, reduced ceramide levels were recorded, suggesting that the ceramide layer is essential for proper barrier function by limiting TEWL. However, it has been argued that the lack of ceramide is due to the high levels of bacteria that colonise atopic skin, especially in chronic AD. Bacterial ceramidases break the ceramides down, resulting in the characteristic low ceramide levels and dry skin seen in AD[46-48]. Interestingly, the enzyme sphingomyelin deacylase, which preferentially metabolises sphingomyelin to sphingosylphosphorylcholine instead of ceramide, is up-regulated in hAD patients[49]. The GCCCGC sphingomyelinase haplotype, furthermore, has been linked with hAD[50], suggesting that a dysfunction in sphingomyelin deacylase and ceramide metabolism could be responsible for the lack of ceramide in hAD and be causative of some of the symptoms.

Two independent loss of function variants in Filaggrin are associated with hAD (R501X and 2282del4)[51;52], this is the strongest evidence to date for a defective skin barrier in AD pathogenesis. Filaggrin is an epidermal barrier protein that contributes to barrier formation by aggregating the keratin cytoskeleton causing the cells to compact into squames. It has also been shown that Filaggrin can be broken down at the stratum corneum to form a 'natural moisturizing substance'[42;53], suggesting that Filaggrin is protective against the characteristic dry skin seen in AD. It has been estimated that Filaggrin mutations occur in as many as 50% of patients with AD[40]. This, however, leaves 50% of patients without Filaggrin as the causative mutation, emphasizing the complex nature of AD and the likelihood that there are multiple causative genes or environmental factors involved in patients with and without normal Filaggrin function.

Other proteins involved in the formation of the stratum corneum are located with Filaggrin on the epidermal differentiation complex (EDC), coded on chromosome 1q21 (in humans). Chromosome 1q21 has been implicated in hAD by linkage analysis[54]. This suggests that other genes involved in epidermal development could also be involved in the pathogenesis of AD.

Allergens could also contribute to the breakdown of the skin barrier. Some allergens are proteases; for example house dust mite allergen Der p 1 is a cysteine protease that has been shown to directly reduce epithelial barrier function[55] and hydrolyse the cell membrane proteins CD23, CD25 and CD40 which potentiates Th2 cell responses[56]. Moreover, exogenous proteases are a potential source of thymic stromal lymphopoietin precursor (TSLP), which can promote allergic inflammation[57].

It is clear that the skin barrier plays an important role in the development of hAD, however evidence for skin barrier dysfunction in cAD is limited[3]. Barrier function studies in dogs have focused on comparing the amount of lipids in plasma, sub-cutaneous fat and skin in atopic and healthy dogs. These studies yielded contrasting results from no difference in levels[58] to significantly lower levels in cAD[59], but these studies looked at the levels of lipids in different ways, at different sites and even at different lipids[60]. Furthermore treatment with essential fatty acid (EFA) supplementation was not associated with consistent changes in plasma or cutaneous EFAs, and showed no correlation with clinical outcome[61]. Nevertheless one study did record, via electron microscopy, that the epidermal lipid barrier was disrupted in canine atopic skin[62]. More recent studies in a colony of atopic Beagles have detected decreased expression of Filaggrin and disturbed extrusion of lamellar bodies by keratinocytes in the epidermis[63]. While these animals do not represent the dog population as a whole, this

does suggests that cAD may display a similar lipid distribution and barrier dysfunction as seen in hAD.

Innate immune response

Once an allergen, irritant, toxin or micro-organism has penetrated the epidermis, it will be recognized by the innate immune response. Innate immunity comprises a series of highly conserved mechanisms thought to have evolved to provide rapid protection against pathogens. This frequently involves recognition of pathogens and other threats using pattern recognition receptors (PRRs)[64]. PRRs are trans-membrane and intra-cellular receptors, e.g. Toll-like receptors (TLR) and the NOD protein family[65]. PRRs identify pathogens via highly conserved molecular patterns such as bacterial cell wall components or peptidogylcan[64]. PRR activation leads to the production of chemokines, cytokines and recruitment of immune cells, as well as determining the extent of the adaptive immune response[56].

It is thought that deficiencies in PRR functionality may affect the maturation of the immune system and lead to high prevalence of atopic diseases and susceptibility of AD sufferers to secondary infection[65]. TLR-binding has been shown to favour the Th2 response[66], and it is possible that the prolonged Th2 response seen in AD could be due to the innate immune system inappropriately responding to innocuous environmental antigens[67].

Keratinocytes express TLRs and also play a role in innate immune function in the skin[67;68]. In hAD, a weak TLR2 response has been observed, and as TLR2 is important in microbial recognition this could explain the increased colonization seen of AD patients[67].

Genetic changes in TLRs or other PRRs could therefore be important in the pathogenesis of AD. Functional changes in PRRs have been shown to affect their protective role in human atopic asthma[69]. A polymorphism in TLR2 has been associated with extreme cases of hAD

with re-current bacterial infections[70], although this conclusion was not supported in another study[71]. TLR9 has been shown to have an association with a certain subset of hAD cases, but TLR6 and TLR4 have shown no association with hAD[72-74]. Nucleotide-binding oligomerization domain-containing protein 1 (NOD1), also known as, Caspase recruitment domain-containing protein 4 (CARD4) is essential in host defence against bacteria and has shown a polymorphism association with hAD[75].

β-definsins and cathelicidins, are effector molecules that form part of the innate immune response. They are antimicrobial peptides which are able to kill microbial pathogens either directly through their interactions with the microbial membrane[76] or indirectly through immunomodulation[77;78]. Because AD patients (humans and dogs) show increased colonisation of *Staphylococcus* and other bacterial species it has been suggested that defects in β-defensin function may be responsible for the development and maintenance of chronic lesional skin in AD. Altered expression of β-defensins has been demonstrated in humans[79;80] and most recently in dogs[81] suggesting a role for β-defensins in the pathogenesis of AD.

Despite growing interest in this field, studies on PRRs and antimicrobial peptides are limited in cAD. However because of the shared pathology and immunological profile, and the fact that the innate immune system is highly conserved, it is thought that evidence from human studies do apply to cAD.

Microbial colonization and infection

As discussed earlier, secondary infection is important in the development and maintenance of chronic lesional AD. A variety of fungal, bacterial and viral infections complicate hAD[48], although only *Staphylococcus* sp. and *Malassezia pachydermatis* are commonly associated with cAD[12]. These can contribute to the pathogenesis and clinical course of the disease in both humans and dogs[82].

Ninety percent of hAD suffers are colonized with *Staphylococcus aureus*, compared to only ~30% of healthy humans[48]. It has been demonstrated in humans that *Staphylococcus* exotoxins can act as super antigens and produce an inflammatory response in the skin[83]. Superantigens are so-called because they elicit a widespread generalized activation of the immune system through cross linking surface MHCII molecules on antigen presenting cells with non-specific T-lymphocytes, rather than the specific binding seen with normal antigens[83]. *S. pseudintermedius* is the main species of Staphylococci found in dogs and has been demonstrated to be present in greater numbers and have greater adherence to keratinocytes in AD compared to healthy dogs and those with other skin diseases[12;84]. No differences in binding ability, susceptibility to beta-defensins, or pulsed field gel electrophoresis (PFGE) types were seen between atopic and healthy dogs, indicating that colonization is associated with host factors[84].

Colonisation and infection with *Malassezia* sp. is also important in human and canine AD; in dogs approximately 50% of *Malasszia* dermatitis cases are concurrent with AD[82]. Numerous human studies have shown that *Malassezia* can act as an allergen in AD[85], and that colonisation modulates cytokine production potentially playing a role in maintaining IgE mediated skin inflammation[86;87]. Specific treatment for *Malassezia* can help ameliorate AD in these patients. Several studies have also demonstrated the presence of *Malassezia* specific serum IgE, positive intra-dermal tests and passive transfer tests in atopic dogs[88-91]. It therefore appears that *Malassezia* can act as an endogenous allergen in cAD, although the clinical significance of this and any response to specific treatment is unknown.

Adaptive immune response

The adaptive immune system is often initiated by the innate responses discussed above. In AD the adaptive immune system responds as illustrated in Figure 1.1 Many genes associated with adaptive immunity have been investigated in AD, as it is likely that some of the changes seen in the adaptive allergic response are due to genetic differences.

The Th2 cytokine cluster genes IL-3, IL-4, IL-5, IL-13, GMCSF (granulocyte-macrophage colony stimulating factor) have been associated with hAD subtypes. Associations have also been demonstrated in RANTES (chemokine (C-C motif) ligand 5) and IL-4R (Interleukin-4 receptor)[38;92], although these were in small populations and subsets of hAD. The similar immunological profile between human and canine AD suggests that similar genetic changes may also be important in dogs.

Hygiene Hypothesis

The inflammatory response seen in AD and other allergic diseases is similar to that seen in helmith infections[93] or ectoparasite infestations[39]. Chronic infection with some parasites leads to a down-regulation of this inflammatory response to protect against inflammation generated tissue damage[93]. It is thought that in diseases such as AD this down-regulatory mechanism does not work effectively, either due to under development or being overwhelmed by the inflammatory response, therefore tolerance is not developed[93-95]. Interestingly it has been demonstrated in dogs that administration of helminths leads to an improvement of cAD symptoms – perhaps 'kick starting' the down regulatory response and inducing tolerance[96].

Observations of a dysfunctional immune response and lack of tolerance to allergens have lead to the 'hygiene hypothesis'. This speculates that improved living standards, reduced

exposure to certain parasitic infections and micro-organisms in early life, affects the development of the immune system. Specifically maturation of cell-mediated (i.e. Th1) immune responses and T-regulatory cell mediated tolerance. In some individuals this could lead to an unregulated Th2 biased immune response and the development of abnormal allergic responses to environmental allergens to which healthy individuals are tolerant[93-95;97]. As described earlier, the rapid increase in AD in industrialised countries in recent years has been considered evidence for the hygiene hypothesis. The exact molecular mechanisms involved in the hygiene hypothesis are not fully understood, however the role of the environment is clear and research in humans is beginning to focus on gene-environment interactions in allergic inflammation. It has already been demonstrated that an individual's genotype can differ according to the microbial products they are allergic to[39]. These findings suggest that specific allergens may differ between individuals according to their genetic background, therefore personalised genetic/medical treatment for AD maybe be important in the future.

Shared environmental factors may affect the occurrence of allergies in both pets and owners. A significant association was found between pet owners with hay fever and the rate of allergy in their dogs[98]. This further illustrates the usefulness of dogs as a model for allergic disease as they share the same environment as humans. There are, however weaknesses in the use of dogs as models for human disease. The presence of multiple breeds and selective breeding of dogs could mean that each breed could have a different genetic cause for the same disease. Secondly the clinical manifestation of disease can be variable in dogs, for example cAD can present has chronic otitis externa, whereas this is not the case in humans, however the use of dogs as models for human disease is of benefit when compared to rodent models.

1.6 Genetic component in canine AD

Most research into cAD has so far been limited to reporting cytokine and chemokine profile changes[31]. Genetic research in cAD is sparse, although there is strong evidence for a genetic link from clinical observations of breed associations with cAD. Specific breeds are more likely to have cAD than others; West Highland white terriers and Boxers are especially susceptible, although many other breeds and crosses are affected[99-101]. Sighthounds, in contrast, are rarely affected; suggesting genetic susceptibility to cAD is not evenly distributed among breeds[102]. However, these studies have rarely taken in account the numbers of these breeds in the whole population, resulting in a skew towards the more popular breeds[101]. Nevertheless, a recent publication from Hungary did take into account the 'over-representation' of certain breeds in the population and found that in most cases their results agreed with previous work[102]. A study comparing multiple geographic regions and veterinary practices reported that breed predispositions to cAD was dependent on the geographic region and even the veterinary practice[103], though this may highlight the popular breeds in those areas rather than a predisposition to cAD. Further evidence for a genetic component in cAD comes from family associations; a study of British guide dogs showed a heritability of 0.47[104]. Heritability is defined as the degree to which a phenotypic trait i.e. disease, is inherited rather than attributable to non-heritable factors i.e. the environment. Stronger evidence for heritability exists in hAD (heritability = 0.72-0.77 in monozygotic twins)[105;106]. Because of the similarities in the immunological profile of cAD to hAD it is thought there is a shared genetic/environmental cause.

Unlike most murine AD models cAD does not appear to be caused by a single gene defect, and, like hAD, may be a polygenic disorder with complex inheritance and interactions with environmental influences. It is therefore a potential complex model for the human

condition. Using dogs as a model to study the genetic basis of AD is advantageous because dog breeds form genetically isolated populations exhibiting strong linkage disequilibrium (LD)[107]. LD occurs when not all the available allele combinations are present in the population, e.g. if loci A and B are represented by two alleles in the population (A and *a*; B and *b*) but only genotypes AABB and *aabb* are found in the population. This is due to the non-random association of alleles at two or more loci. The degree of LD describes how likely it is that these alleles are inherited together. The degree of LD in dogs is extensive, meaning that fewer genetic markers are required to find an association with disease in dogs and smaller sample sizes can be used to find these associations, compared to human genetic studies[107;108].

An association study or a linkage study can therefore be undertaken to investigate the genetics of cAD. When trying to characterise the genetics of cAD the heterogeneity of the breeds, the possibility of different disease expression and variable diagnosis, and the difficulty of making direct comparisons between breeds must all be considered[109].

1.7 Approaches to genetic study of complex diseases and findings in AD

Linkage studies

Two or more loci and their representative genes are said to be linked if they occur on the same chromosome. Contrary to the second of Mendal's laws of inheritance, eukaryotic genes which are linked will tend to be inherited together and are not randomly assorted. Cross-over and/or chromosome translocation can prevent linked loci from being inherited together, if cross-over and/or translocation do not take place this can lead to LD. If a loci is linked with a disease phenotype it will be in LD when disease affected and unaffected individuals in the same family are compared in a linkage study. Multiple families are usually analysed in order to find a common linked loci, which if found may suggest that the disease loci was inherited

from a common ancestor and could be disease causative in the general population. This approach is relatively unbiased but is limited to diseases where large relative risks are conferred by a limited number of genes. This is because for disease causing genes with a relative risk <2, large numbers are required to detect linkage. One study for example illustrated that 2500 families are required to detect a relative risk of 2[110].

In dogs, however, the situation is different, as they form genetically isolated populations which exhibit strong LD[107;108], and 46% of heritable dog diseases are breed specific[111]. Therefore, it can be said that certain breeds 'are enriched for a small number of disease alleles, which are rare in the overall dog population'[108]. The implications are that linkage analysis can be carried out using dramatically reduced numbers over an extended canine family, making this powerful approach viable for complex disease studies.

Although no linkage studies in cAD have been reported, there have been a considerable number of linkage studies on hAD (Table 1.1). However human studies may fail to identify all or some genes which are relevant for cAD. Moreover, linkage studies only provide an area of linkage and not the specific gene or polymorphism associated with the disease. Nevertheless areas of linkage can serve as a starting point for candidate gene studies.

Table 1.1 Areas of Linkage Identified in Human AD

Chromosome location	Marker	Phenotype	Population & size	Reference
11q13	D11S97	Extrinsic AD	64 nuclear families	112
13q12-14	D13S218	AD	German AD = 192; German controls = 59; German parents of AD = 77; Sweden: 40 nuclear families	113
5q31-33	D5S436 – D5S643	AD		
3q21	D3S3606	AD	Germany: 199 sib pairs	114
1q21	D1S498	AD		
17q25	D17S784	AD	UK: AD = 153, Controls = 230 (sib pairs)	54
20p	D20S115	AD and asthma		
3q24-22	D18S851	AD	Sweden: 109 nuclear families	115
3q14	D3S1768	SCORAD*		
13q14	D3S2459	SCORAD*		
15q14-15	D1751290	SCORAD*		
17q21	D15S118	SCORAD*		
18q21	D13S325	Extrinsic AD		
3p26-24	D3S3594-D3S3038	Extrinsic AD	Denmark: 100 nuclear families	116
18q11-12	D18S877	Extrinsic AD	Denmark: 100 nuclear families	
2p12	D2S1777	Increased IgE		
16q21	D16S2620	AD and asthma	Multiethnic European: 82 nuclear families	117
3q21.3	D3S3606	Extrinsic AD		

*SCORAD - SCORing Atopic Dermatitis Index, severity score for human AD

Candidate gene/association studies

Association studies are usually the standard approach for investigating complex diseases. These are case-control comparisons that endeavour to identify polymorphisms (e.g. Single nucleotide polymorphisms) in candidate genes. The basis of this type of study being; if a polymorphisms frequency significantly differs between cases and controls it may be implicated in the disease pathogenesis. This type of study is more sensitive to small multiple gene effects[110], but it is often biased and limited as it is largely restricted to small regions of genes that have already been annotated or previously linked with the disease. Again, the extensive LD exhibited by dog breeds allows for smaller sample sizes to be used to detect genetic risk factors and increase the statistical power.

Currently no candidate gene studies have been undertaken for cAD. Genes identified by candidate gene studies in hAD are shown in Table 1.2.

Table 1.2 Genes Identified from Human AD Candidate Gene Studies

Gene Name	Gene symbol	Chromosome Location	Population & size	Reference
Filaggrin	FLG	1q21	Irish n=52, Scottish n=279, Danish n=142	52
Cytotoxic T lymphocyte associated 4	CTLA4	2q33	Australian: 112 nuclear families	118
Toll-like receptor 9	TLR9	3p21.3	British n=172	119
Interferon regulatory factor 2	IRF2	4q35.1	Japanese: 48 nuclear families (n=180)	120
Granulocyte-macrophage colony-stimulating factor	GM-CF	5q31.3	British n= 113	121
Interleukin-13	IL-13	5q31-33	Japanese n=185	122
			German n=187	123
Interleukin-4	IL-4	5q31-33	German n=187	123
			Japanese: 88 nuclear families	124
Serine protease inhibitor kazal type 5	SPINK5	5q31-33	British n=338	125
			Japanese: 41 nuclear families (n=177)	126
			Japanese n=124	127
Caspase recruitment domain-containing protein 4	CARD4 (NOD1)	7p14-15	German: 189 nuclear families (n=454)	75
B chain of the high affinity receptor for IgE	FcεRIβ	11q12-13	British: 60 nuclear families (n=277)	128
Interleukin 18	IL-18	11q22	German n=225	129
T cell immunoglobulin and mucin domain containing molecule 1	TIM1	12q12-13	Korean n=112	130
PHD finger protein 11	PHF11	13q14	Australian: 111 nuclear families	131

Table 1.2 continued

Mast cell chymase	CMA1	14q11.2	Japanese n=100	132
			Japanese n=145	133
			Japanese n=169	134
			British 341 nuclear families	135
				136
			German n=242	
IL-4 receptor alpha chain	IL4R	16p12-p11	Japanese n=27	137
			Japanese n=101	138
Caspase recruitment domain-containing protein 15	CARD15 (NOD2)	16q12	German n=330	139
Regulated on activation normally T cell expressed and secreted	RANTES	17q11.2	German n=268	140
Eotaxin	EOTAXIN	17q21.1-21.2	Japanese n=140	141
TGF-β1	TGF-β1	19q13.1	British n=68	142
Stratum corneum chymotryptic enzyme	SCCE	19q13.3	British n=103	143
Glutathione S-transferase, Theta 1	GSTT1	22q11.2	Russian n=325	144
BCL2-related protein A1	BCL2A1	15q24.3	Caucasian: AD=105, Controls=110	145
Late cornified envelope-like praline-rich 1	LELP1	1q21	North India: 133 nuclear families & independent case-control cohort	146
interleukin 1 receptor-like 1	IL1RL1	2q12	Japanese: AD=452, Control=636	147
Sphingomyelinase	SMPD2	11	Korean: AD(intrinsic)=251, AD(extrinsic)=284, Controls= 188	50

Genome Wide Association Studies

Ideally, an unbiased approach with extensive coverage should be used to investigate gene associations with disease. These criteria are fulfilled by genome-wide association studies (GWAS)[148]. These studies use known Single Nucleotide Polymorphisms (SNPs) spaced throughout the genome, and compare case and control frequencies of these SNPs. This identifies genes and SNPs that may be causative of the disease phenotype. The advantage of this approach is that it is unbiased as it looks at the whole genome. Nevertheless, it is still limited to known SNPs, and this approach can be expensive depending on the number of SNPs and samples used. To achieve whole genome coverage in human studies it is estimated that 200,000 to 500,000 SNPs are required[108], dramatically increasing the cost of this approach. In dogs, however, the extensive LD exhibited by dog breeds[107] means that only 5,000 to 30,000 SNP markers need to be used, making genome wide association a viable option for canine research[108].

In GWAS and other studies measuring frequencies of SNPs, Hardy-Weinberg equilibrium (HWE) is an important consideration when interpreting the results. HWE states that allele frequencies in a population remain in equilibrium, from generation to generation, unless there is non-random mating, mutations, selection, limited population size, random genetic drift and gene flow. In every population one or more of these effects is likely to be taking place, but nevertheless, HWE is used in human genetic studies as a quality check for genotyping data by comparing observed genotype frequencies to those expected within a population. The statistical tests used to measure genetic associations within a population are based on HWE which means SNPs must be within HWE to be associated with the disease. It is thought that dog breeds may deviate from HWE, because they are selectively bred (non-random mating). Nevertheless it has been shown that when dogs are grouped into pedigrees (breeds) and HWE

is measured they do conform to HWE more than if grouped as multiple breeds[149], this is probably due to different breeding histories. This means that when interpreting SNP frequency data, HWE must be calculated in order to exclude SNPs which deviate from HWE and are therefore unreliable. By analyzing one breed at a time the likelihood that SNPs will conform to HWE is increased[149].

mRNA expression studies

Microarrays can be used to identify disease associated genes by comparing specific gene expression (mRNA) profiles in affected and unaffected individuals. If significant dysregulation of a gene is observed in affected individuals it is possible that the gene is involved in the disease pathogenesis, and is therefore a good candidate for further study. Microarrays have the ability to examine a large number of genes, covering most tissue transcripts. One problem with this approach is that any observed dysregulation may be due to downstream inflammatory effects rather than aetiological cause. Furthermore, microarrays are prone to producing false positive results, although statistical corrections such as the Bonferroni correction[150] or the Benjamini and Hochberg correction[151] reduce the number of false positives. Microarray results should be validated by qPCR (preferably in larger sample sizes), as this technique is more sensitive and allows absolute quantification of the amount of transcript. As microarrays use RNA samples extracted from the diseased and control tissue, unlike the techniques described earlier, there are no known advantages to using dogs, apart from the benefits of confirming dysregulation of the same genes in a different organism.

1.8 Hypothesis

The working hypothesis for this project is that cAD is a complex disease with a significant genetic component, contributing to the pathogenesis.

1.9 Aims of the project

The aims of this project were to: (1) Investigate the genetic factors involved in the pathogenesis of cAD, and (2) to identify specific gene associations with cAD within and between dog breeds.

To do this a dual approach of candidate gene association study and GWAS was used. This was considered to give a novel unbiased approach and also use information from previous studies on which to base gene/SNP selection. This combined approach was taken in order to increase the likelihood that any genetic factors in cAD would be identified. Figure 1.3 shows the overall study design for this project.

Figure 1.3 Study Design

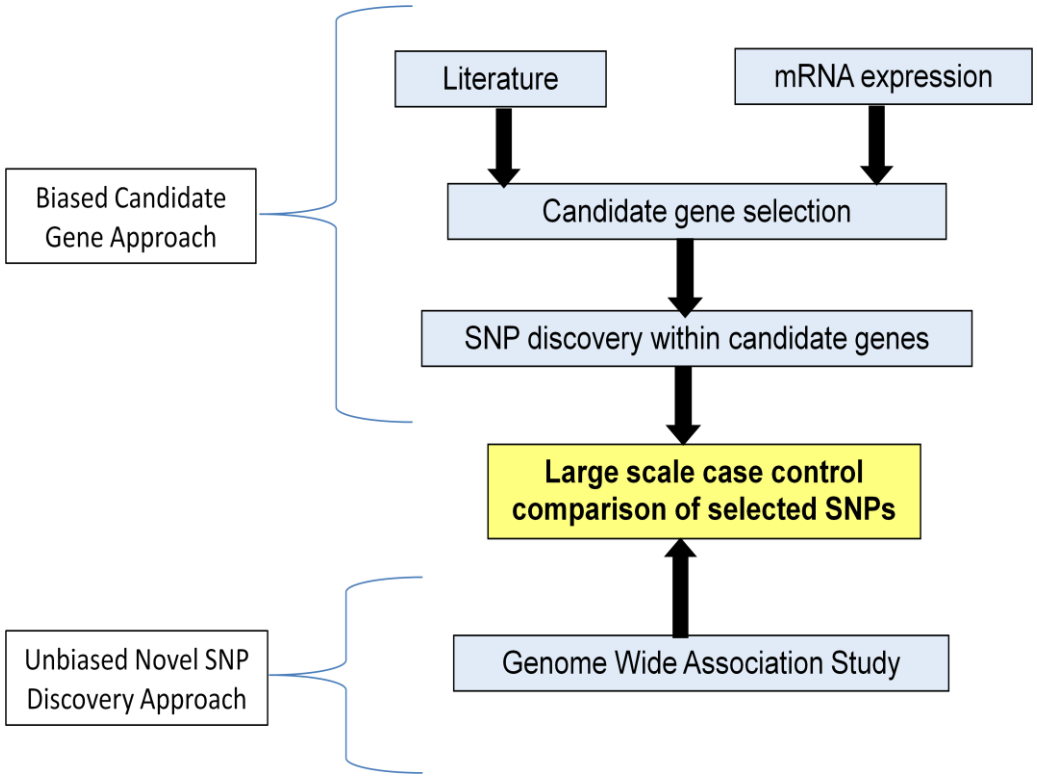


Figure 1.3: A dual approach of candidate gene and genome wide association study (GWAS) to select SNPs for a large-scale case-control comparison to test associations with cAD. The biased approach: Candidate genes were selected from the literature and by analysis of an mRNA microarray study. The unbiased approach: GWAS

Chapter 2

General Materials & Methods

2. General Materials & Methods

2.1 Case and control selection

Diagnosis

Dogs with cAD were recruited from the dermatology clinic at the University of Liverpool Small Animal Teaching Hospital and through the WALTHAM® Centre for Pet Nutrition. Clinical diagnosis of cAD was based on compatible history and clinical signs, and exclusion of other causes of pruritus. Coat brushings, skin scrapings, *Sarcoptes* specific IgG serology and trial therapy were used to eliminate the possibility of an ectoparasite infestation. All atopic dogs underwent a six to eight week diet trial with either a home-cooked or commercial single, novel protein diet, or a commercial hydrolysed protein diet and water only, to eliminate the possibility of an adverse food reaction. Seborrhoea, staphylococcal pyoderma and *Malassezia* infections were managed appropriately. No anti-inflammatory medication was given for at least three weeks prior to examination.

All dogs with a clinical diagnosis of AD, from the University of Liverpool Small Animal Teaching Hospital, had an intra-dermal allergen test (IDT) with 54 environmental allergens (Greer Laboratories Inc., Lenoir, NC, USA), performed and interpreted according to accepted criteria[152]. Clinical lesions were scored using the Canine Atopic Dermatitis and Severity Index (CADESI) -03. This is a validated assessment of clinical lesions (erythema, excoriation, lichenification and self-induced alopecia) at 62 anatomical sites from 0 (normal) to 5 (most severe) yielding a score of 0-1240[153].

Control dog DNA samples were obtained from the Veterinary Laboratories agency (VLA), Bath Championship Dog show and WALTHAM® Centre for Pet Nutrition. Control animals from the dog show and WALTHAM® Centre for Pet Nutrition did not have any

history or show any clinical signs compatible with cAD, Otis externa, immune regulated or skin diseases. The VLA samples represented a random population with no disease information.

Various breeds were sampled and specific information on breed, sex and age is given in subsequent chapters.

Sample collection

Blood

All blood samples were taken by a veterinarian from the jugular vein using a 1 inch 21 gauge needle. Approximately 1-3 ml was taken and placed in an EDTA tube to prevent clotting. Blood was taken for diagnostic purposes, and only blood which was excess to clinical requirements was used for this study. The blood samples were stored at -40°C and used with informed consent from the owners of each dog. The study followed ethical guidelines as laid down by The University of Liverpool, The University of Manchester, and the Biology and Biotechnology Research Council.

Skin

Atopic skin samples were taken from dogs undergoing a diagnostic procedure as part of their treatment plan. Where necessary, the dogs were sedated with 10-20mg/kg medetomidine (Dormitor®; Pfizer Animal Health, Sandwich, UK) given IV. 0.5-1.0ml of 2% lignocaine without adrenalin was then infiltrated into the subcutis under the sites marked for biopsy. Non-lesional samples were taken from clinically unaffected skin on the flank. Lesional samples were taken from areas of erythema and macular-papular dermatitis from the ventral body. Grossly evident excoriation, staphylococcal pyoderma and *Malassezia* dermatitis were avoided.

Control samples were taken from healthy dogs that had been euthanized for reasons unconnected with the study, with no history and clinical signs of pruritus or conditions likely to alter immune function. There was a delay of approximately 30 minutes between euthanasia and sample collection.

In both cases six millimetre diameter skin biopsies were taken and either snap frozen using liquid nitrogen or placed into RNAlater™ (Ambion Inc., Austin, TX, USA). RNAlater™ and snap frozen samples were stored at -80 C.

The tissue samples were excess to clinical requirements, stored and used with informed consent from the owners of each dog. The study followed ethical guidelines as laid down by The University of Liverpool, The University of Manchester and the Biology and Biotechnology Research Council.

Buccal Swabs

Buccal swabs were taken from control dogs attending the Bath Championship Dog show. A foam applicator was rubbed on the inside of the dogs' cheeks and then applied to the Whatman Indicating FTA® cards. The FTA® cards were used because they allowed storage of the buccal samples at room temperature.

The buccal swabs were stored and used with informed consent from the owners of each dog.

2.2 Sample processing

DNA extraction from blood

A previous experiment carried out by myself and others[154] indicated that Qiagen Qiamp DNA blood midi kits (Crawley, UK) provide the best quality DNA for the blood samples used in this study. The DNA was extracted according to the manufacturer's instructions and assessed for quality using the Nanodrop® spectrophotometer ND-1000 (Thermo scientific, Delaware, USA). The DNA was then normalized to 20ng/µl or 50ng/ µl with distilled water or 10xTE using the Hamliton Microlab® Star (Bonaduz, Switzerland). The DNA was then stored at -80°C or -20ºC until it was required.

DNA extraction from buccal FTA® cards

Two 6mm punches were taken from the FTA® cards (Whatman Inc., Florham Park, NJ) and the DNA was extracted using a variation on a previously optimised method[155] . Figure 2.1 shows this optimized method.

Figure 2.1 DNA extraction process from buccal FTA cards (Lema *etal* (2006)[155])

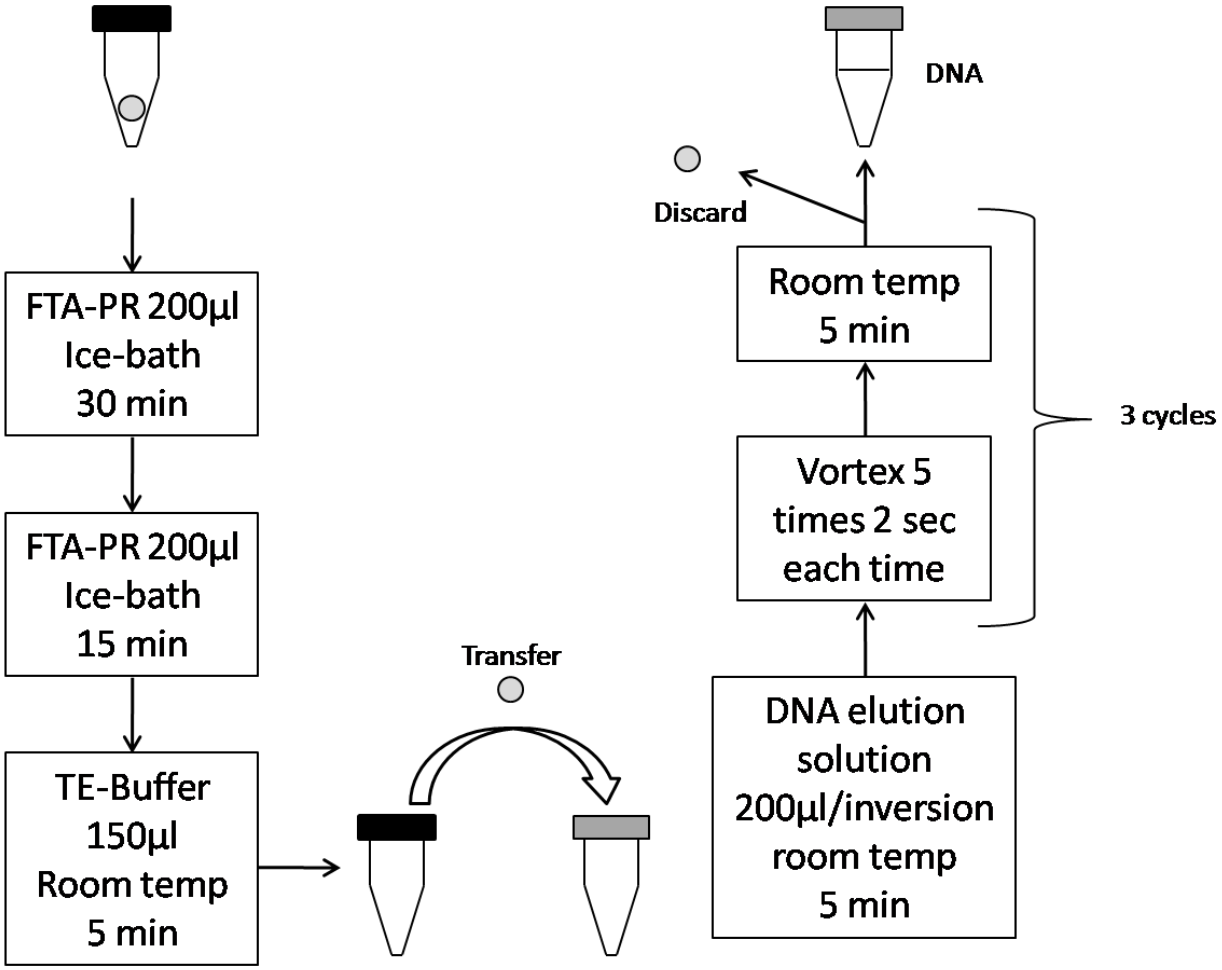


Figure 2.1 shows the method for DNA extraction from FTA cards. FTA-PR is FTA® Purification Reagent (Whatman Inc., Florham Park, NJ). A 6mm punch of the FTA card was placed into an eppendorf with 200µl FTA-PR and incubated in an ice bath for 30 minutes. The FTA-PR is discarded and a further 200µl was added, the sample was kept on ice for a further 15 minutes. This was discarded and 150µl of TE buffer (10ml 1M TRIS, 0.2ml 0.5M EDTA up to 1 litre with distilled water) were added and kept at room temperature for 5 minutes. The punch was transferred into a clean eppendorf where 200µl of DNA elution solution (100ml 1M NaOH, 100ml 1M TRISHCL, 0.625ml 0.5M EDTA up to 1 litre with distilled water) were added, the tube was inverted at room temperature for 5 minutes. The sample was then votexed 5 times for 2 seconds and left to stand at room temperature for 5 minutes. This was repeated 3 times. The punch was discarded leaving a DNA solution.

The protocol above was modified by doubling the volumes of FTA® Purification Reagent (Whatman Inc., Florham Park, NJ) and TE-1 buffer to allow for 2 punches being used instead of one. The volume of DNA elution solution (100ml 1M NaOH, 100ml 1M Tris-HCL, 0.625ml 0.5M EDTA, bring up to 1L) was unaltered; this was to maximize the concentration of DNA extracted. The DNA was stored at -20ºC until required.

2.3 RNA extraction from skin, quality assessment & cDNA synthesis

Total RNA was extracted from the skin biopsies using the dismembranator (Mikro-Dismembrator, Sartorius Stedim Biotech, France). The powder was then placed in Trizol® (Sigma, Poole, UK) and extracted using a chloroform alcohol technique and a RNA extraction kit from Qiagen (Crawley, UK). The integrity and quality of the RNA was checked using the Agilent 2100 Bioanalyzer (Agilent Biotechnologies Inc., Santa Clara, CA, USA).

cDNA synthesis was performed via reverse transcription using Superscript II reverse transcriptase (Invitrogen, Paisley, UK) according to the manufacturer's instructions (http://www.invitrogen.com). Initially 200µg (10µl) total RNA were pre-incubated with 0.5µg (1µl) oligo-dT (Invitrogen) and 10mM (1µl) dNTP mix (Invitrogen) at 65°C for 5 minutes. 4µl of 5x first strand buffer (containing 250 mM Tris-HCl (pH 8.3), 375 mM KC1, 15 mM $MgCl_2$), 2µl of 0.1M DTT and 40 units (1µl) of RNase inhibitor (Promega, Southampton, UK) were added to each sample and the samples incubated for 2 minutes at 42°C, followed by the addition of 200 units (1µl) of Superscript II reverse transcriptase (Invitrogen) and incubation for 50 minutes. Reverse transcriptase activity was terminated by incubation at 70°C for 15 minutes.

2.4 Laboratory techniques

The techniques described here are presented as an outline to the methodology; individual differences to the methods will be presented in subsequent chapters.

Quality assessment of DNA

The concentration and 230/240 and 260/280 ratios were measured using the Nanodrop® spectrophotometer ND-1000 (Thermo scientific, Delaware, USA). The 260-280nm ratio of samples needed to fall within 1.8 to 2 to be uncontaminated and be used in this study.

Quality assessment of RNA

The Bioanalyser 2100 (Agilent Biotechnologies Inc., Santa Clara, CA, USA) was used to assess the quality of the RNA using a RNA 6000 Nano Chip Kit (Agilent Biotechnologies). A RNA Integrity Number (RIN) is returned. In this study samples with a RIN score lower than 6 were excluded.

Normalisation, dilutions and sample sorting

The Hamilton robot was used to normalize the DNA to either a concentration of 50ng/µl or 20ng/µl. All DNA was stored in the UK DNA Companion Animals archive at either -40ºC or -80ºC. The XL20 robot was used to cherry pick samples for the study using 2D barcodes on each sample tube. Finally the Cybio robot was used in some PCR protocols to dilute DNA or to add SAP solution to PCR products.

RNA Microarray

The canine microarray was designed and preformed by WALTHAM® Centre for Pet Nutrition; details of the assay are summarised in Chapter 3.

Q PCR assay design

The qPCR assays were all performed in triplicate using a TaqMan™ ABI PRISM 7900 SDS (Applied Biosystems, Foster City, CA, USA) in a 384-well plate format. A 10-µl reaction volume was used per well. This consisted of 5 µl Taqman 2X PCR master mix (Universal PCR Mastermix; Applied Biosystems), 0.1 µl each of 20 µM forward and reverse primers (MWG Biotech, Eurofins, Germany) 0.1 µl of 10 µM probe (Exiqon; Roche Diagnostics Ltd), 0.1 µl distilled water and 4.6 µl of sample cDNA or water for the negative controls.

The amplification was performed according to the standard ABI 7900 protocol: 10 minutes at 50°C followed by 40 cycles of 95°C for 1 minute and 60°C for 15 seconds, as recommended by the manufacturer (Applied Biosystems). The real-time data were analysed by Sequence Detection Systems software, version 2.2.1 (Applied Biosystems). The detection threshold was set manually at 0.05 for all assays. Standard curves were generated for each assay, to confirm the efficiency of the assay is between 93%-107% and the R^2 value was > 0.98.

Touchdown PCR

Touchdown PCR was used as opposed to traditional PCR because it reduces non-specific priming and therefore the background 'noise' caused by non-specific DNA fragments is reduced. It does this by gradually lowering the annealing temperature as the PCR cycles progress. Initially the annealing temperature is set 7°C above the temperature of melting (T_m) of the primers used, at each cycle the temperature is reduced by 0.5°C. The initial higher temperature increases specificity, and the subsequent lower temperatures permit more efficient amplification from the specific products formed during the initial cycles.

The touchdown PCRs were performed throughout to optimize primers and were on either a DNA engine Dyad™ peltier thermal cycler or a DNA engine Tetrad™ PTC-225 peltier thermal cycler (MJ research Inc., Waltham MA, USA) using these cycling conditions: 95ºC for 5minutes, followed by 95ºC for 30seconds, Ta+7ºC for 30 seconds (reduce by 0.5ºC per cycle) and 72ºC for 1minute for 13 cycles, and 95ºC for 20 seconds, Ta for 1 minute, and 72ºC for 1 minute for 19 cycles, and finally 72ºC for 5minutes.

Gradient PCR

Gradient PCR was used when touchdown PCR failed to optimise primers, this allowed a temperature range to be run across the primers establishing the correct temperature for amplification. Once this temperature was established touchdown PCRs were performed.

Electrophoresis-gel quality check

All PCR products were run on a 2% agarose gel with a 1kb ladder (Promega, Southampton, UK) to assess the concentration and size of the PCR product.

Transgenomic WAVE

Details of sample selection, primer design, cycling conditions and equilibration are in Chapter 6. Conditions were generated for each fragment using WAVEMAKER V4.1.44 (Transgenomic, Huston, Texas, USA) software.

Microsatellite genotyping

The initial PCR assays were all performed on the Dyad™ peltier thermal cycler or a DNA engine Tetrad™ PTC-225 peltier thermal cycler (MJ research Inc., Waltham MA, USA) in a 384-well plate format. A 10µl reaction volume was used per well: 0.04µl Hot start Taq (5U/µl Qiagen, Crawley, UK), 0.05µl each of 50pmol/µl forward (labelled) and reverse

primers (Metabion, Germany), 0.3µl $MgCl_2$ (25mM, Qiagen, Crawley, UK), 0.5µl DNTPs (2mM), 1µl x10 buffer (Qiagen, Crawley, UK) 3.06µl distilled water and 5µl of DNA at 5ng/µl. The cycling conditions were as follows: 95ºC for 5 minutes, then 94ºC for 20 seconds, 55ºC for 30 seconds and 72ºC for 30 seconds for 10 cycles, 89ºC for 20 seconds, 55ºC for 30 seconds and 72ºC for 30 seconds for 20 cycles and finally 72ºC for 10 minutes. ABI prism® Rox400 (Applied biosystems, Foster City, CA, USA) was used for accurate sizing. The ABI prism® 3100 genetic analyzer data collection software V1.1 (Applied biosystems, Foster City, CA, USA) was used to set up the run. The ABI prism® Genescan software V3.7 (Applied biosystems, Foster City, CA, USA) was then used to analyse the data. ABI prism® Genotyper software V3.7 (Applied biosystems, Foster City, CA, USA) was used to group the results into genotypes/microsatellites. Full details are in Chapter 6.

Illumina canine SNP20 GWAS array

The Illumina canine SNP20 GWAS array was performed by the Barts and the London Genome Centre according to the manufacturers' instructions; assay details are in Chapter 7.

Sequenom genotyping

Sequences of genes to be genotyped were retrieved from Ensembl using the Biomart data mining tool[156] and exported in FASTA format. A PERL script was used to convert the FASTA file into a suitable format for the Sequenom design tools. Using ProxSNP and Prextend[157] quality control procedures were run on the sequences applied, rejecting unsuitable sequences for the Sequenom platform. The SNPs were then plexed together by the assay design software (MassARRAY) in a way that means the greatest chance of genotyping success and the largest number of SNP possible was included on each plex.

All primers and probes were supplied by Metabion (Germany); the probes were diluted using distilled water to 400µM and the primers to 100µM using the Hamilton Robot. The primers were then pooled together in plexes, each primer within the pool at a concentration of 500nM done by the Hamilton robot. Probes were also pooled into plexes although the plex was split into low mass and high mass probes; low mass at a concentration of 7µM and high mass at 14µM, again performed on the Hamilton robot. The primers and probes were then ready for PCR, details of which are given in Chapters 8 & 9.

Chapter 3

Candidate Genes & Microarray analysis of Gene Expression

3. Candidate genes & Microarray Analysis of Gene Expression

The mRNA microarray analysis has previously been published: Merryman-Simpson A.E., **Wood S.H.**, Fretwell N., Jones P.G., McLaren W.M., McEwan N.A., Clements D.N., Carter S.D., Ollier W.E., Nuttall T.J. (2008), Gene (mRNA) expression in canine atopic dermatitis: microarray analysis. *Veterinary Dermatology*, 19(2):59-66.

3.1 Introduction

The literature review suggested that cAD has a heritable component with many complex interactions. It was hoped that by using a candidate gene approach and investigating gene expression and polymorphisms, the pathogenetic mechanisms contributing to disease aetiopathogenesis would be identified.

In this chapter I will consider literature-based candidate gene selection and mRNA microarray analysis of gene expression.

Candidate genes are usually selected for their identified function suggested in exsisting literature. For example AD is often associated with keratinocyte hyper-proliferation; it is therefore logical to select a gene that is responsible for keratinocyte proliferation. Though previous association or linkage based studies also prove useful in gene selection. Literature searches are inevitably biased by personal opinion and current trends in research.

Microarray analysis of gene expression (mRNA microarrays) allows comparison between disease-associated and control gene expression by measuring a selection of mRNA transcripts on a microarray chip. Microarray chips have multiple thousands of known RNA sequences attached, each representing an individual gene/protein. Figure 3.1 shows the preparation of an mRNA microarray chip.

Figure 3.1 mRNA Microarray preparation and analysis

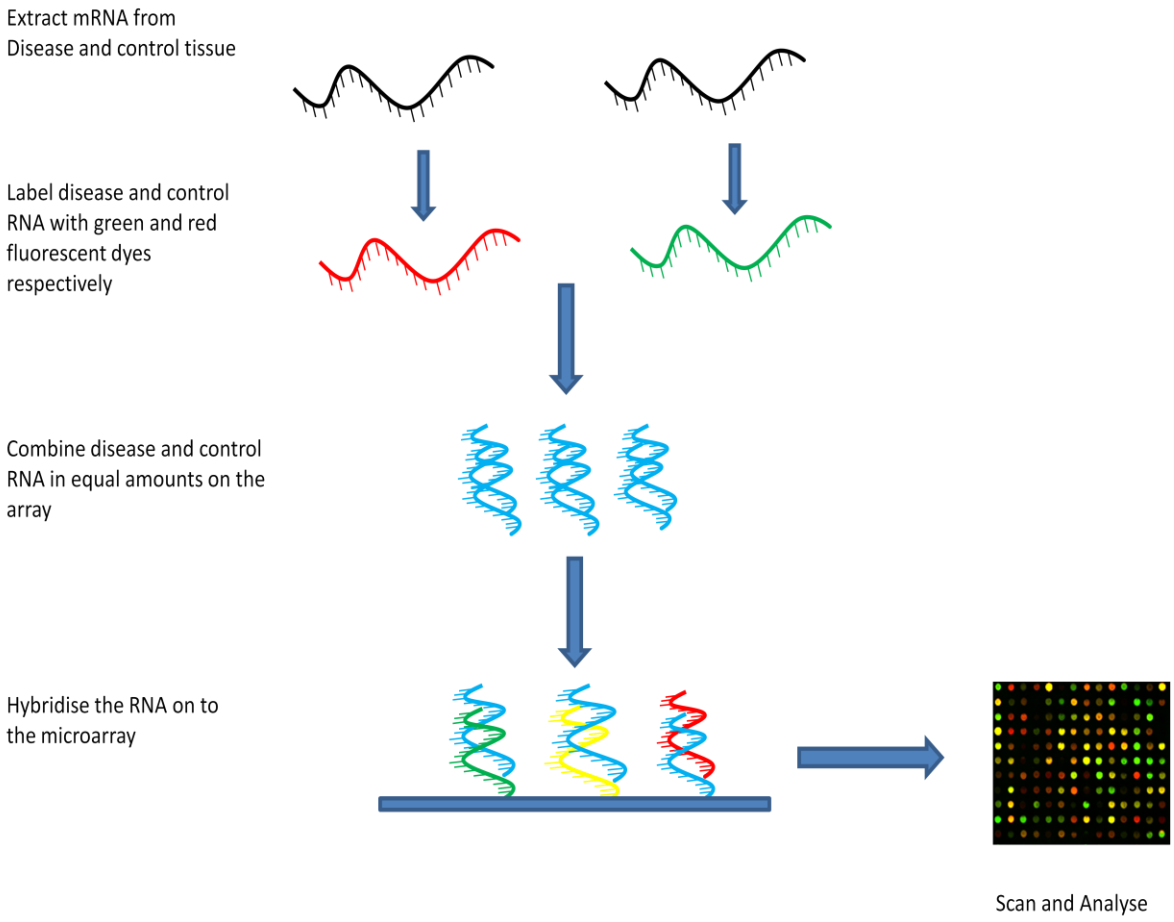


Figure 3.1: RNA is extracted from control and diseased tissue samples. The RNA is then amplified using a PCR reaction. The control samples are mixed with red dye and the diseased with green dye. The control and disease samples are hybridized to the microarray and by measuring the intensity of dyes on the microarray it is possible to identify genes that are up-regulated, unchanged and down-regulated in each (i.e. if a transcript is up-regulated in the disease state the intensity of the green dye will be greater than the red for that transcript and *vice versa*).

This method allows for an unbiased approach in candidate gene selection as the transcripts used cover a large portion of the genome, including hypothetical proteins and genes of unknown function. This broad sweep approach can potentially identify novel genes in the pathogenesis of cAD. Literature searches can be useful in narrowing down genes identified by microarray analysis by inferring function.

mRNA microarray analysis measures the levels of expression of mRNA transcripts for large numbers of genes and can justifiably be used in a relatively small number of case and control samples, keeping costs down whilst still providing a wealth of data. Statistical correction of mRNA microarray data is required as mRNA microarrays are prone to false positive results[158]. However, this correction can lead to crucial genes being omitted from the analysis.

The purpose of this study was to identify potential candidate genes involved in cAD using literature searches and a canine mRNA microarray. To do this Ensembl[159;160] and Pubmed[161] were used for the literature searches and an mRNA microarray designed by WALTHAM® Centre for Pet Nutrition and supplied by Agilent with 22,000 transcripts was utilized. Control skin from healthy dogs was compared to two disease skin samples (lesional and non-lesional) from each atopic dog. It was hoped the literature search and mRNA microarray would supply good candidate genes for further genetic study in cAD.

3.2 Methods

Case recruitment and sample collection

Diagnosis and sample collection/storage was performed as detailed in Chapter 2. This study compared 17 cAD cases with nine healthy controls. The cases included 12 pure breeds (Labrador retriever, German shepherd, Rhodesian ridgeback, West Highland white terrier, Staffordshire bull terrier, Scottish terrier, Jack Russell terrier, bulldog, Boxer, bull mastiff, Springer spaniel, Neapolitan mastiff) and five cross breeds. Seven were male (13 months – 8 years 1 month) and 10 were female (7 months – 9 years 2 months). The healthy control dogs included seven pure breeds (Jack Russell, German shepherd, Staffordshire bull terrier, Labrador, Boxer and English bull terrier) and two cross breeds. All controls were male dogs (18 months - 8 years), All the animals were privately owned.

RNA extraction and quality assessment

This was performed as detailed in Chapter 2.

Agilent 22K oligonucleotide canine array

The canine microarray experiments were designed and performed by WALTHAM® Centre for Pet Nutrition. All the data was analysed by me.

The array was designed by definition of coding fragments from the 1.2x poodle genome sequence in collaboration with Celera (Celera diagnostics, Rockville, MD[162]). An in-house 54,000 protein database and the NCBI reference sequence database[163] were used to isolate matching islands for dog coding sequences using the Basic Local Alignment Search Tool (BLAST)[164]. Non-redundant sequences matching the NCBI reference sequence database (cut-off of e-40) and the Protein database (cut-off of e-20) were submitted for oligonucleotide design to Agilent technologies (Agilent Technologies, Palo Alto, CA). Three oligonucleotides per sequence were designed and tested against mRNA derived from canine blood, testes and cell culture. Based on maximum performance on the chip and uniqueness in the gene set, one of the three oligonucleotides was selected.

Following the first round of *in vitro* transcription the amplified RNA underwent a second round of amplification and fluorescent labelling with Cyanine 5-CTP (PerkinElmer, Bucks, UK) or Cyanine 3-CTP (PerkinElmer, Bucks, UK) using the low RNA input fluorescent linear amplification kit protocol (Agilent Technologies, Palo Alto, CA). The Cyanine labelled amplified RNA samples were subsequently purified using RNeasy® spin columns (Qiagen Ltd, West Sussex, UK) and were eluted in a volume of 60µl.

Target solutions for the microarray chips were produced by combining 8 pmol of Cyanine-3 and Cyanine-5 labelled cRNA and 1x control targets in 1x fragmentation buffer as

described in the Agilent 60-mer oligo microarray processing protocol. The hybridisation mix was loaded onto the 22K microarray which was incubated for 17 hours at 60^{o}C with rotation at 4 rpm.

The microarrays were initially washed in 6 x SSPE (0.75M NaCl, 50mM NaH_2PO4, 5mM EDTA, pH 7.0) + 0.005% N-lauroylsarcosine followed by a 0.06 x SSPE+ 0.005% N-lauroylsarcosine wash and finally a wash in acetonitrile Far UV (Fisher Scientific, Leic, UK). The microarrays were scanned at a resolution of 5µm in an Agilent G2565AA microarray scanner system (G2567AA Feature extraction software, version 7.1, Agilent Technologies, Palo Alto, CA.). The images were analysed using an Agilent Technologies software package.

Statistical analysis of microarray data

Data were imported into Genedata Expressionist Analyst (Genedata AG, Basel, Switzerland), and the Cy3 and Cy5 fluorescence intensities normalised using lowest weighted linear regression (LOWESS). Normalised data in tab delineated text format were uploaded into NIA array analysis[165]. This allowed a pair-wise comparison of gene signal intensity (expression) with fold difference of two or greater and statistical significance of $P \leq 0.05$, as determined by single factor analysis of variance with multiple hypothesis correction by the false discovery rate (FDR). The fold change was calculated for each sample to represent a ratio of expression between disease and control samples. This number was further corrected to provide the actual fold change in the sample as follows: ratio -1 if the sample was up-regulated and by (-1/ratio) +1 for down-regulated samples. This is stated as the true fold change.

Literature selected candidate genes

PubMed[161] and Ensembl[159;160] databases containing research articles and gene sequence information, were employed to research genes and/or chromosomal regions with reported associations with cAD and/or hAD. These databases were also used to identify proteins with functions relevant to the disease pathogenesis.

3.3 Results

Microarray analysis

Fifty four genes out of 22,000 were found to be significantly differentially expressed in cAD skin when compared to healthy control skin ($p=<0.05$). Sixteen of the 54 genes showed an increase or decrease in expression in both lesional skin and non-lesional skin from dogs with AD when compared to skin from healthy controls (Table 3.1, Figure 3.2). Twelve of the 54 genes were only differentially expressed in lesional skin when compared to healthy controls (Table 3.2, Figure 3.3), and the remaining 26 were only differentially expressed in non-lesional skin (Table 3.3, Figure 3.4). Only one gene, S100 calcium binding protein A8 (S100A8), showed a significant increase in expression when comparing lesional to non-lesional skin (Table 3.2). The true fold change in expression between healthy, lesional and non-lesional skin and statistical significance are presented in Tables 3.1, 3.2 and 3.3, with the genes grouped according to function. Figures are presented here to visually represent the patterns of expression (Figures 3.2, 3.3 and 3.4).

Table 3.1 Genes Differentially Expressed in Both Lesional and Non-lesional Atopic Skin Compared to Skin from Healthy Controls Grouped According to Function

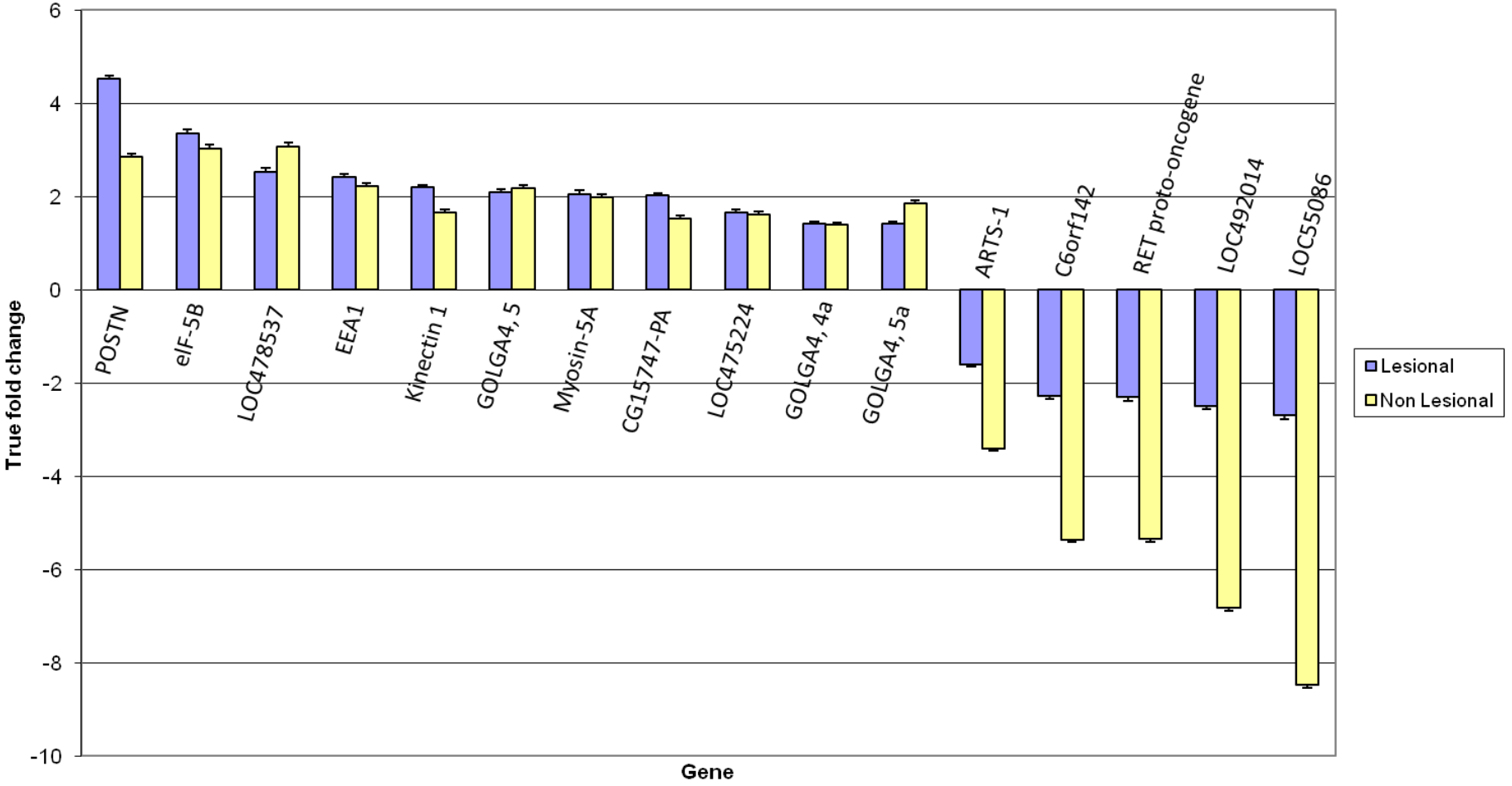
		Lesional skin		Non-lesional skin	
Gene name	**Accession number**	**True fold change***	**P value corrected by (FDR)§**	**True fold change***	**P value corrected by (FDR)§**
Inflammation/Immunology					
LOC488898 (ARTS-1)	XP_546015.1	-1.60	0.00	-3.4	0.00
Cell cycle/apoptosis/repair/lesion formation					
LOC477298 (POSTN)	XP_534490.1	4.52	0.00	2.85	0.00
LOC478537 (down-regulated in ovarian cancer 1 isoform 2)	XP_535715.2	2.53	0.04	3.07	0.01
Translational control					
LOC474556 (eIF-5B)	XP_531784.1	3.36	0.00	3.03	0.00
Transport/regulation					
LOC480332 (kinectin 1)	XP_537455.1	2.19	0.00	1.66	0.01
LOC478312 (Myosin Va)	XP_535487.1	2.05	0.02	1.98	0.02
LOC475224 (A-kinase anchor protein 9 isoform 2)	XP_532456.1	1.65	0.05	1.61	0.04
Canis familiaris ret proto-oncogene (RET)	XP_543915.1	-2.31	0.03	-5.3	0.00
Barrier formation					
Golgin subfamily a, 5	CAA54261.1	2.09	0.00	2.19	0.00
Golgin subfamily a, 4	AAC51791.1	1.41	0.02	1.39	0.01
Golgin subfamily a, 5	NP_002069.2	1.41	0.04	1.85	0.00
Miscellaneous					
LOC475424 (EEA1)	XP_532649.1	2.42	0.00	2.23	0.00
LOC610752 (CG15747-PA)	XP_868363.1	2.02	0.00	1.54	0.01

Table 3.1 continued

C6orf142 (chromosome 6 open reading frame 142)	NP_612636.1	-2.28	0.00	-5.4	0.00
LOC492014 (sushi-repeat-4)	XP_549134.1	-2.50	0.00	-6.8	0.00
Hypothetical protein LOC55086	AAH70110.1	-2.70	0.03	-8.5	0.00

∗ True fold change represents the actual change in expression. §FDR is the false discovery rate which was used to correct the P values, all of which are <0.05. The results show five significantly down-regulated genes and 11 significantly up-regulated genes.

Figure 3.2 True Fold Change of Gene Expression in Lesional and Non-lesional Skin Compared to Control Skin*



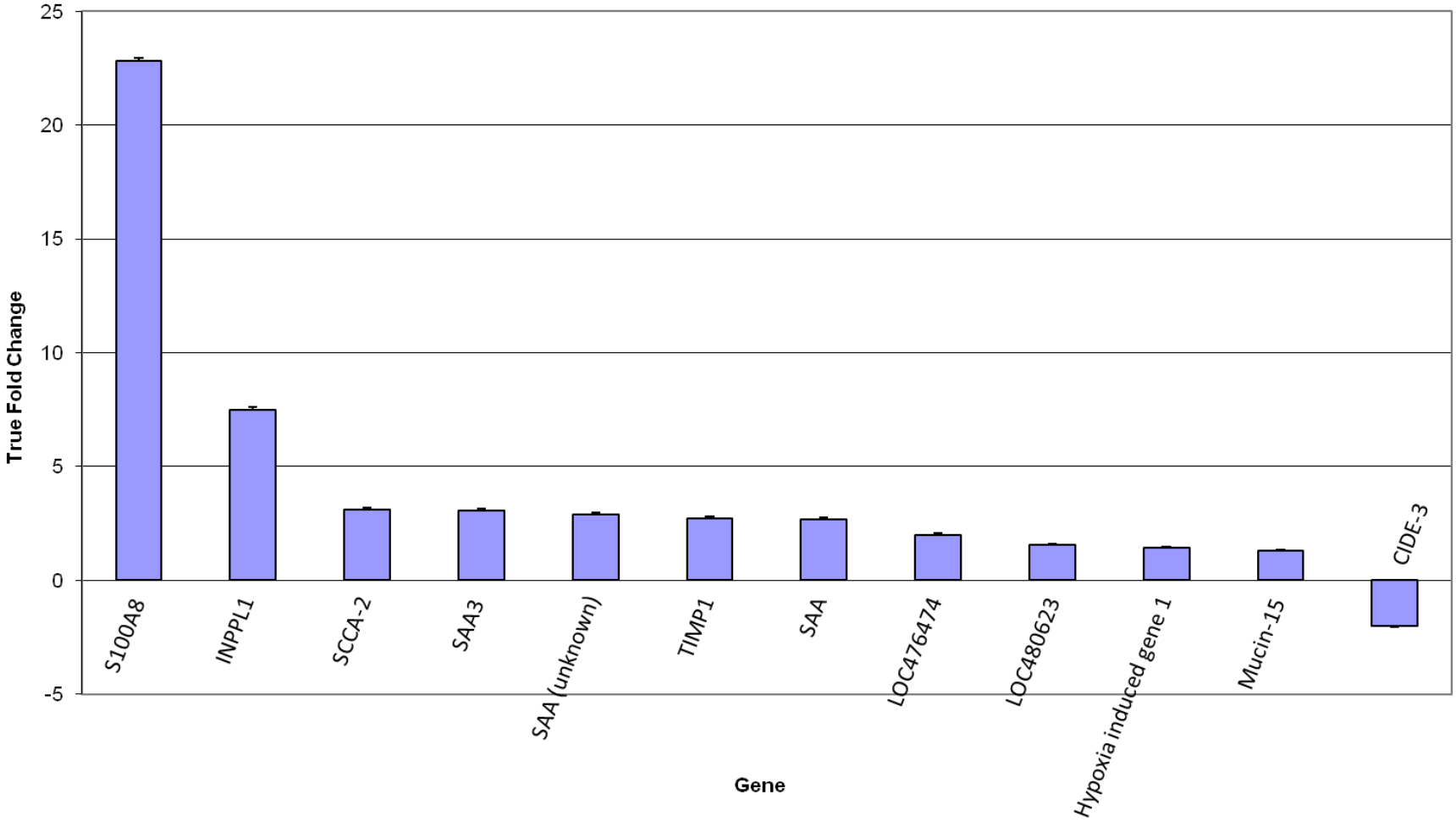
*The error bars represent standard error

Table 3.2 Genes Differentially Expressed Only in Lesional Skin when Compared to Healthy Controls.

Gene name	Accession number	True fold change*	P value corrected by (FDR)§
Inflammation/Immunology			
LOC490461 (S100A8)* *	XP_547583.1	22.81	0.00
LOC 485209 (INPPL1)	NP_001558.2	7.49	0.00
LOC483954 (SCCA-2)	XP_541074.1	3.11	0.00
LOC480749 (SAA3)	XP_537868.1	3.07	0.01
serum amyloid A protein [Canis familiaris]- no records	AAA62765.1	2.91	0.00
LOC403816 (TIMP1)	BAA32393.1	2.72	0.04
LOC751814 (SAA)	AAA62762.1	2.67	0.01
Cell cycle/apoptosis/repair/lesion formation			
LOC474674 (similar to RAD50 homolog isoform 1)	XP_531901.1	2.00	0.04
LOC484662 (CIDE-3)	XP_541777.1	-1.99	0.00
Transport/regulation			
LOC480623 (Sperm-associated antigen 5 (Astrin))	XP_537743.1	1.56	0.05
Barrier Formation			
LOC143662 (Mucin-15)	AAH20912.2	1.29	0.05
Miscellaneous			
LOC477037 (hypoxia induced gene 1)	XP_534235.1	1.42	0.05

* True fold change represents the actual change in expression. §FDR is the false discovery rate which was used to correct the P values, all of which are <0.05. The results show one significantly down-regulated gene and 11 significantly up-regulated genes. ** Represents significant difference between lesional and non-lesional skin.

Figure 3.3 True Fold Change of Gene Expression in Lesional Skin Compared to Control Skin*



*The error bars represent standard error

Table 3.3 Genes Differentially Expressed Only in Non-lesional Skin when Compared to Healthy Controls

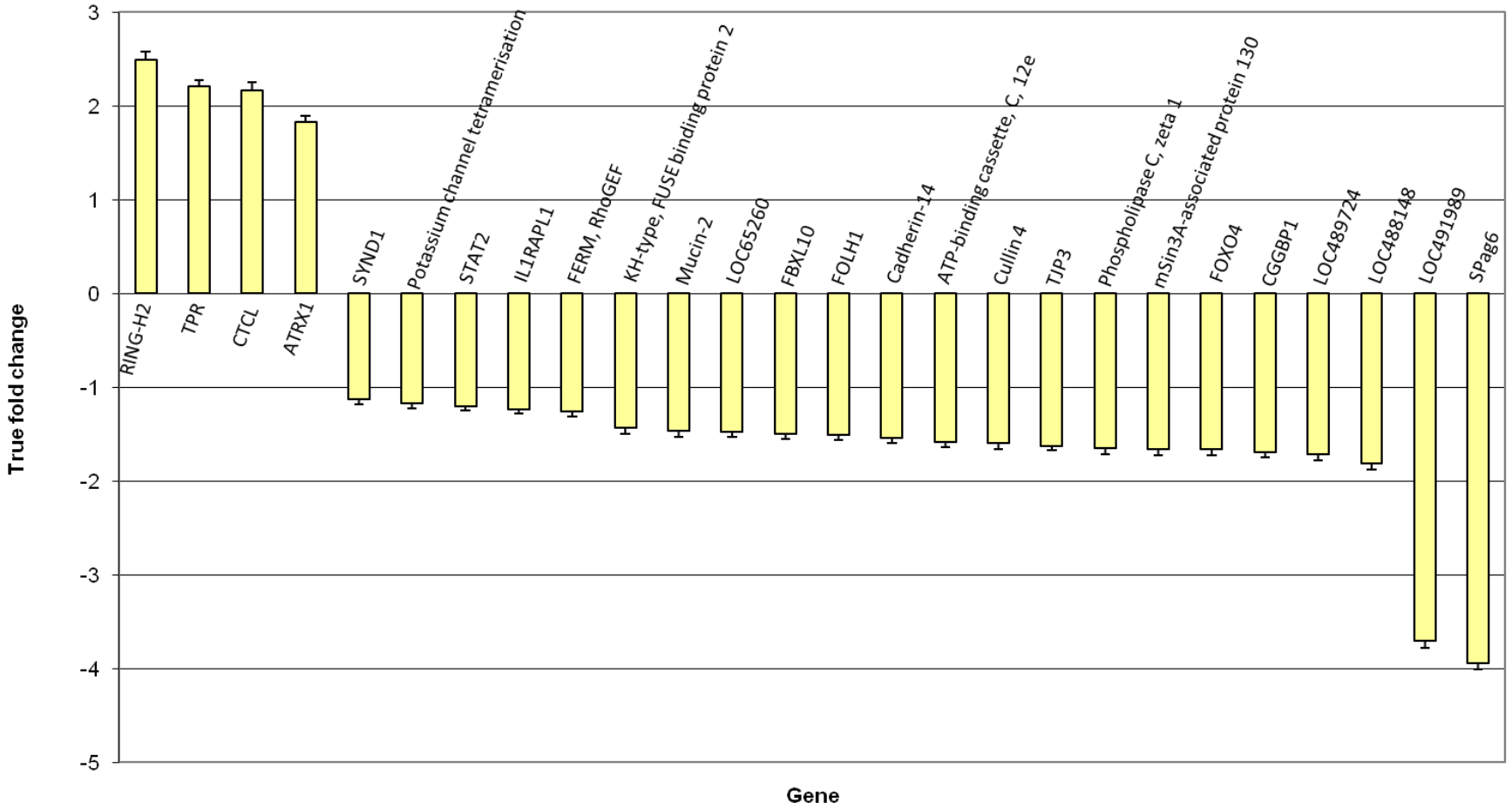
Gene name	Accession number	True fold change*	P value corrected by (FDR)§
Inflammation/immunology			
C10orf118 (CTCL (cutaneous T cell lymphoma) tumor antigen L14-2)	AAM44457.1	2.16	0.03
IL1RAPL1 (interleukin 1 receptor accessory protein-like 1)	NP_001009038.1	-1.24	0.03
Cell cycle/apoptosis/repair/lesion formation			
LOC482986 (SYND1)	XP_540099.1	-1.13	0.04
LOC489256 (Cadherin-14)	XP_546374.1	-1.54	0.00
LOC8451 (Cullin 4A)	AAR13072.1	-1.60	0.03
Transport/regulation			
LOC480045 (Nucleoprotein TPR)	XP_537167.1	2.21	0.05
LOC479689 (potassium channel tetramerisation)	XP_536820.2	-1.18	0.04
LOC486205 (FERM, RhoGEF and pleckstrin domain protein 2)	XP_543330.1	-1.26	0.01
LOC487294 (ATP-binding cassette C12e)	XP_544420.1	-1.58	0.01
TJP3 (tight junction protein 3)	NP_001003202.1	-1.63	0.00
LOC486657 (phospholipase C, zeta 1)	XP_543784.1	-1.65	0.05
Barrier formation			
LOC483653 (Mucin-2)	XP_540774.1	-1.47	0.03
Transcription factor			
LOC481111 (STAT2)	XP_538232.1	-1.21	0.02
CGGBP1(CGG triplet repeat binding protein 1)	AAD04161.1	-1.70	0.00
LOC485022 (FUSE binding protein 2)	XP_542140.1	-1.44	0.03
FOXO4 (foxhead box)	XP_549066.1	-1.66	0.03

Table 3.3 continued

Miscellaneous			
LOC477332 (RING-H2 protein)	XP_534526.1	2.49	0.03
LOC480963 (ATRX1)	XP_538084.1	1.83	0.01
C1orf163 (hypothetical protein LOC65260)	AAH15313.1	-1.47	0.02
LOC 465184 (FBXL10)	XP_520652.2	-1.50	0.01
LOC2346 (FOLH1)	AAC83972.1	-1.51	0.02
LOC491989 (S6K-alpha 6)	XP_549109.1	-3.71	0.00
LOC476102 (mSin3A-associated protein 130)	XP_533311.1	-1.66	0.02
LOC489724 (PH domain leucine-rich repeat protein phosphatase-like)	XP_546844.1	-1.72	0.02
LOC488148 (similar to ecotropic viral integration site 1)	XP_545272.1	-1.82	0.01
LOC50525 (Spag6)	NP_056588.1	-3.94	0.00

∗ True fold change represents the actual change in expression. §FDR is the false discovery rate which was used to correct the P values, all of which are <0.05. The results show 22 significantly down-regulated and four significantly up-regulated genes.

Figure 3.4 True Fold Change of Gene Expression in Non-lesional Skin Compared to Control Skin



*The error bars represent standard error

Literature & Microarray selected candidate genes

The candidate genes selected for further study are summarised in Table 3.4. Candidate genes were chosen based on relevant function, any previous associations with hAD and/or significant changes in expression on the canine microarray.

Table 3.4 Summary of Potential Candidate Genes in cAD, selected from Literature Searches and mRNA Microarray

Gene Name	Evidence From:	Supporting References
Barrier Function		
GOLGA4	Canine microarray	45;166
LOC483653 (homologue to MUC2)	Canine microarray	167
Filaggrin (FLG)	Functionality, association, microarray and linkage.	51;52;168;169
SPINK5	Function and association	126;170
SGPL1	Function and RT-PCR	45;49;171
CSTA (cystatin A)	Functionality and microarray	45;172;173
SPRR1B (cornifin)	Function and microarray	169;174;175
PKP2 (plakophillin)	Function and microarray	176-180
CARD4	Function and association	45;75
Immunology/Inflammation		
TSLP	Function	45;181;182
INPPL1 (SHIP2)	Canine microarray	183
LOC480749 (homologue to serum amyloid SAA1)	Canine microarray	184;185
LOC483954 (homologue: SCCA2 or SERPINB4)	Canine microarray	186
LOC490461 (homologue:S100A8)	Canine microarray	187
LOC488898 (homologue:ARTS-1)	Canine microarray	188

Table 3.4 continued

IL1RL1	Function and association	147;189
CTLA4	Function and association	118
DPP4	Function and microarray	178;190-193
MS4A2	Linkage and function	43;45;112
CMA1	Linkage and function	43;194
Cell cycle/apoptosis/repair/lesion formation		
LOC477298 (homologue:POSTN)	Canine microarray	195
TIMP1	Canine microarray	196
CUL4A	Canine microarray	197
LOC482986 (SYND1 – CD138 antigen)	Canine microarray	198
SOCS3 (Suppressor of cytokine signalling 3)	Linkage, microarray and function	115;199;200
CDH13 (cadherin 13)	Function and microarray	178;201;202
Calcium Transport		
LOC486657 (phospholipase C, zeta 1)	Canine microarray	203
Transcription Factor/misc.		
LOC481111 (homologue: STAT2)	Canine microarray	204
PPARγ/α	Function and microarray	178;205-209

3.4 Discussion

This study set out to identify potential candidate genes involved in disease pathogenesis. The literature searches suggested 16 genes with either functions relevant to AD pathogenesis or reported associations to hAD. The canine mRNA microarray identified 54 genes that were dysregulated in cAD suggesting that a number of complex interactions lead to the atopic phenotype. A number of these genes have functions that potentially link them with cAD. Table 3.4 summarises the candidate genes selected for further study and the evidence upon which the selection was based, these included: function, human AD skin mRNA microarray studies, canine skin mRNA microarrays, qPCR, genetic linkage and association. The results of the microarray analysis are discussed below; the results of the literature search will be also be discussed in this section, under the sub-heading "literature selected candidates" (p.62).

mRNA canine Microarray

The greatest increase in gene expression observed in this study was by S100 calcium binding protein A8 (S100A8), which showed a 22.81 fold increase in lesional skin compared to control skin. The S100 gene family encode pro-inflammatory molecules released by phagocytes in inflammatory conditions[187] and their levels have been shown to 'track' the severity of inflammatory diseases[210]. Release is stimulated by tumour necrosis factor (TNF)-α, a cytokine expressed in both canine and human AD[187]. The RAGE/NF-κβ pathway is thought to maintain release of TNF-α and subsequent release of S100[211] via a positive feedback loop. Previous studies have shown that TNF-α expression is also correlated with the degree of inflammation in both canine and human AD[1;31]. S100A8 has also been linked to leukocyte recruitment, suggesting that it is a key gene in the inflammatory response[187]. In this study, S100A8 was not significantly over-expressed in non-lesional skin, suggesting that it has a purely inflammatory role in lesional skin.

Serum amyloid A (SAA) was over-expressed in lesional skin, and is produced in response to inflammation. SAA is an apolipoprotein that recruits immune cells to inflammatory sites and is regulated by interleukin (IL)-1, IL-6 and TNF-α[184;185].

Squamous cell carcinoma antigen 2 (SCCA2) is another gene associated with inflammation and is induced in response to IL-4 and IL-13. SCCA2 has been shown to disrupt the enzyme function of *Dermatophagoides* house dust mite allergens which are known triggers of allergic responses[186], suggesting SCCA2 may be released as a protective measure. The dogs included in this study all had positive intra-dermal test reactions to *Dermatophagoides farinae* and *D. pteronyssinus*. SCCA2 expression, however, was only increased in lesional skin, indicating that expression will have little impact on early responses to dust mite allergen exposure.

Several other genes differentially expressed only in lesional skin (Table 3.2, figure 3.3) were mainly inflammatory genes and, with the exception of cell death-inducing DFF45-like effector-3 (CIDE3), were over-expressed. Genes expressed exclusively in non-lesional skin, in contrast, were generally under-expressed (Table 3.3, Figure 3.4) and represent a range of functions. The majority of these down-regulated genes are involved in the transport of calcium, potassium and other ions. Calcium transport is increased in inflammatory conditions and calcium is essential for the function of some pro-inflammatory molecules. The down-regulation of calcium transport may represent a lack of an anti-inflammatory stimulus in non-lesional skin. Differences in non-lesional skin compared to healthy control skin were unexpected; suggesting that non-lesional skin in cAD is not 'normal' and could represent "pre-lesional" skin in AD.

Increased levels of Signal transducer and activator of transcription 2 (STAT2), have been linked to allergic driven inflammation[212] therefore, it would be expected that lesional skin

would show increased STAT2 expression when compared to control skin but no dysregulation of STAT2 was observed in lesional skin. However, STAT2 expression was decreased in non-lesional skin compared to control skin. Altered expression of STATs has also been linked with a number of inflammatory barrier-related diseases such as irritable bowel syndrome (IBS) and psoriasis[213;214], and with human atopic diseases such as rhinitis and asthma[215;216], suggesting that there may be a role for STATs in cAD. Furthermore, STAT is part of the JAK-STAT pathway[217] which is the way interleukins transduce the signals that initiate cytokine specific gene transcription. Whilst only STAT2 was included in this array, there are another six STAT genes within the family that perform a range of functions, some antagonistic to each other. It would therefore have been interesting to include all the STAT genes on this microarray.

Forkhead box O4 (FOXO4), a transcription factor, was down-regulated in non-lesional skin. Forkhead box P3 (FOXP3) is a related gene that is expressed on human T-regulatory cells[218], whose expression leads to inhibition of the immune response[219]. Defects in T-regulatory cells are associated with immune pathologies; for example FOXP3 deficient mice develop allergic airway inflammation and exaggerated IgE responses[220]. If FOXO4 performs similar functions, initial down-regulation may lead to the poor T-regulatory cell function hypothesised in cAD[31], resulting in increased levels of IgE, chronic inflammation and lesion formation. However, the function of FOXO4 has not yet been defined in any species.

Not all human and canine AD cases present with increased IgE levels[1] indicating that there is more to AD than allergic hypersensitivity. It has been suggested that skin barrier formation may play a role in the development of human and canine AD[42;43]. A recent study suggested that differences in barrier function could be involved in both IgE and non-IgE

associated hAD (see Figure 1.2 for details)[44]. Genes relating to skin barrier/epidermis formation could, therefore, be important in the pathogenesis of cAD.

Strong genetic linkage to hAD has been previously reported within the epidermal differentiation complex (EDC) on chromosome 1q21[44;110;221]. The EDC comprises genes such as pro-Filaggrin, loricrin, involucrin, S100 and small proline rich proteins (SPRPs), which are essential for keratinocyte differentiation and epidermal barrier formation[221;222]. A large-scale microarray study of hAD[169] found S100 to be up-regulated in lesional skin, as observed in this study, whilst Filaggrin and loricrin were down-regulated. We were not able to study Filaggrin and Lorcrin using the present canine array, as at the time of designing the array the importance of these genes was unknown. Moreover, designing primers for Filaggrin is difficult due to its repetitive sequence. With the exception of a few microsatellite markers I am currenlty unaware of any group that has successfully designed primers for canine Filaggrin.

Ceramide is a lipid essential for barrier function and formation; it covalently binds to involucrin, envoplakin, elafin, SPRPs and loricrin to form a lipid layer and the cornified envelope[223]. Ceramide maintains the skin moisture levels by limiting trans-epidermal water loss (TEWL)[45]. In human atopic patients; atopic skin has significantly lower amounts of ceramide compared to healthy skin, and the TEWL in atopic patients is 2 fold greater in non-lesional skin and 4 fold greater in lesional skin when compared to healthy skin[45]. It has also been demonstrated by chromatography that ceramide levels in non-lesional cAD skin decreased[224]. Ceramide was not on the canine array used but the ceramide transporter Golgi auto-antigen Golgi subfamily a4 (GOLGA4) gene was present[166]. The study results showed that GOLGA4 was up-regulated in lesional and non-lesional atopic skin suggesting that although GOLGA4 associated ceramide transport is enhanced, ceramide levels are still low in

atopic skin. The reported low ceramide levels may be associated with post-transcriptional defects or other functional abnormalities. Increased bacterial colonisation of atopic patients may explain the decreased levels of ceramide seen in atopic skin[46]. Bacterial ceramidases break down ceramide, giving the characteristically low ceramide levels and dry skin seen in atopic dermatitis[46]. This is consistent with the hypothesis that bacterial colonisation is linked to chronic AD[47;48]. It is possible that the increased expression of GOLGA4 is in response to low cutaneous ceramide levels.

Both human and canine AD are characterised by spongiosis of the epidermis. This is most severe at lesional sites, though present at non-lesional sites[225]. Syndecan-1 (SYND1) is present in the keratinocyte membrane where it modulates cellular adhesion, regulating the cell architecture[226]. Down-regulation of SYND1 is associated with spongiosis[227]. In this study, SYND-1 was down-regulated in non-lesional skin but not in lesional skin. The reason for this is unclear although other inflammatory factors may influence expression. For example in IBS[198] increased levels of TNF-α cause down-regulation of SYND1, suggesting that if the levels of TNF-α were increased at non-lesional sites relative to lesional sites, this may have caused the down-regulation observed. As discussed in Chapter 1, AD shows a biphasic T cell response and is characterized by the presence of TNF-α in the early stages and IFN-γ in the latter stages. This suggests that non-lesional skin is perhaps "pre-lesional" skin (i.e. that is skin which is atopic but is yet to show clinical symptoms) and may have increased TNF-α levels which lead to the decrease SYND1 levels discussed above. This reasoning is however speculative, as TNF-α expression was not measured in this study.

Chronic lesions in human and canine AD are characterised by fibrosis and lichenification. This is promoted by periostin (POSTN) and tissue inhibitor of metalloproteinase 1 (TIMP1)[195;196]. These genes may therefore contribute to lesion formation in AD. On the

canine microarray used, POSTN was over-expressed in both lesional (not statistically significant) and non-lesional skin (statistically significant). TIMP1, in contrast, was only over-expressed in lesional skin. In human studies, serum levels of TIMP1 were greater in atopic patients and levels correlated with the chronicity of lesions and degree of lichenification[196].

This study has identified genes expressed in atopic lesional and non-lesional canine skin compared to normal healthy canine skin. The identification and inclusion of IgE responders and non-IgE responders as discrete groups (i.e. atopic dermatitis and atopic-like dermatitis) on the array may have further explained the pathogenic mechanisms of cAD. Given the previous implication in hAD of Filaggrin and involucrin, inclusion of these genes on the array would have been beneficial.

Microarray experiments provide a rapid, wide ranging, unbiased assessment of many genes, allowing novel genes to be identified in the pathogenesis of cAD. However, it is relatively insensitive to small changes in expression which, when dealing with a complex disease such as cAD could result in important genes with small effects being missed. Given the incidence of false positive results when using microarray[158], the Benjamini and Hochberg correction[151] was used. This gives a false discovery rate (FDR) and represents the probability the findings were due to chance. In this study, all FDRs were between 0 and 0.05 indicating the findings in this study were robust. Nevertheless, the results will need to be confirmed using quantitative PCR (qPCR); a more sensitive, accurate and quantitative method.

Literature selected candidate genes

Candidate genes selected from literature searches are presented in Table 3.4, which is divided into functionally relevant sections. These sections will be used to structure the discussion of the candidate genes.

Barrier formation

The stratum corneum[221] is formed from the basal layer and involves synthesis of keratins and cornified envelope proteins in the keratinocytes. The cornified envelope forms following interactions between involucrin, envoplakin, elafin, small proline rich proteins (SPRPs) and loricrin. Ceramide covalently bonds to these proteins to form a lipid layer[223]. Human genetic linkage studies demonstrate that chromosome 1q21, where these genes and Filaggrin are located, is linked with hAD[168].

Filaggrin is down-regulated in hAD in human microarrays[169]. Since the canine array was designed, multiple studies have reported the importance of Filaggrin in hAD. Two loss of function variants were identified in filaggrin[52] and other mutations and associations have been found[51;228-231]. The variants R501X and 2282del4 were discovered in nine families. Linkage analysis confirmed that they are predisposing factors for AD, leading to the conclusion that AD is inherited as a 'semi-dominant trait with high penetrance in homozygotes' and a 'reduced penetrance in heterozygotes'[52]. Filaggrin also contributes to barrier formation by causing the cells to compact into squames by aggregating to the keratin cytoskeleton. Filaggrin can be broken down at the stratum corneum to form a 'natural moisturizing substance'[42;53]. A decrease in the amount of Filaggrin, as seen in hAD, could lead to dry skin and decreased barrier function. Because of the importance of Filaggrin in hAD it is an excellent candidate for cAD research. More recent studies have revealed that atopic canine skin features higher levels of trans-epidermal water loss, decreased ceramide levels[224] and reduced expression of Filaggrin when compared with healthy canine skin[63]. However a decrease in Filaggrin expression has only been observed in an experimental colony of IgE sensitized Beagles. These Beagles are an experimental model for cAD but do

not represent the clinical disease in the dog population as a whole, therefore Filaggrin levels in clinical cases of AD across different breeds remains to be quantified.

SPRR1B also represents a good candidate for further research, being a component of the cornified envelope, it is located in the same region as Filaggrin and is involved in epidermal formation via keratinocyte differentiation and peptide cross linking[174;175]. Furthermore, it has been shown to be up-regulated in a human microarray study[169].

PKP2 is a structural component of desmosomes[176]. Desmosomes are specialized cell junctions which are essential for adhesion and tensile strength of epithelia. There are many sub-components to desmosomes such as cadherins, desmogleins, desoplakins etc[177]. Because of their role in maintaining skin barrier strength it is possible that these could be involved in the pathogenesis of cAD. Furthermore, PKP2 has been shown to be down-regulated in microarray experiments[178] in hAD. PKP2 expression has been implicated in beta catenin/T cell factor signalling activity[179] and T-cell survival[180], further suggesting a possible role in cAD.

SGPL1 is part of a complex pathway involved in lipid metabolism, and a key modulator of ceramide metabolism[232], the importance of which in AD was discussed earlier. Its expression was shown to be increased in a human real-time PCR study[171].

Cystatin A produces a precursor of one of the proteins in the cornified envelope in keratinocytes, and it also plays a role in epidermal development and maintenance[233;234]. Cystatin A is also a cysteine protease inhibitor of house dust mite protease[235], which has shown an association with hAD[236].

CARD4 (NOD1) has been associated, in a German human population, with AD[75]. It encodes cytosolic pathogen recognition receptors, with a preference toward recognition of

gram-negative bacteria, and because of this it has been suggested that CARD4 may represent a 'novel barrier-based pathogenic mechanism underlying the development, maintenance, or both of AD'[43].

SPINK5 functions in hair and skin morphology, proteolysis during epithelial formation and keratinocyte differentiation[170]. This suggests that altered SPINK5 expression will have implications for skin barrier function[237]. SPINK5 mutations are associated with the atopic phenotype in both British[125] and Japanese[126;127] human AD cohorts. It is also associated with Netherton syndrome in humans, where two of the symptoms are atopy and ichthyosis[125;238].

Immunology and inflammation

Thymic stromal lymphopoietin (TSLP) is an IL-7-like cytokine which has been implicated in allergic disease[181;182]. Immunohistochemistry in humans has shown that there is increased expression of TSLP by keratinocytes in atopic skin lesions in hAD[239]. Mouse studies have also suggested a role for TSLP in the pathogenesis of AD; over-expression of TSLP in the mouse model epidermis leads to the development of AD symptoms and cytokine profile[240;241]. TSLP has been referred to as the 'allergic master switch'[182] for allergic inflammation. Observations of TSLP-treated human blood dendritic cells, demonstrate an induction of an inflammatory Th2 cytokine pattern[239]. A recent study has demonstrated that TSLP also 'converts human epidermal Langerhans cells into antigen-presenting cells that induce pro-allergic T cells'[242]. Previous studies have noted that Langerhans cells serve as antigen-presenting cells in AD[243;244], and expression of high-affinity IgE receptors allows for more efficient allergen capture, processing and presentation to T cells[245]. This effect of TSLP could render Langerhans cells capable of inducing or maintaining AD symptoms[242].

IL-1 receptor-like 1 (IL1RL1) is a protein coding gene for an IL receptor. Studies of the gene in mice suggest that IL1RL1 can be induced by pro-inflammatory stimuli, and it is also

thought to be involved in helper T cell function[246]. IL1RL1 has many splice variants, and one of these, ST2L, has been linked to hAD because it is expressed in Th2 but not Th1 cells[189]. A more recent study identified a SNP in the distal promoter of IL1RL1 gene that is significantly associated with human AD. Moreover, this SNP directly affects the expression level of ST2L[147].

Cytotoxic T-lymphocyte-associated protein 4 (CTLA4) is in the immunoglobulin superfamily and encodes a protein which transmits an inhibitory signal to T cells. Altered expression of CTLA4 has been implicated in autoimmune disease. Although there is little to support a potential role for CTLA4 in AD, one study found an association with human (infant) AD and CTLA4 polymorphisms[118].

Membrane-spanning 4-domains, subfamily A, member 2 (Fc fragment of IgE, high affinity I, receptor for; beta polypeptide; MS4A2) has been shown to be in an area of linkage with human AD. MS4A2 produces FcεRIβ which is important in IgE function[43;112], as discussed in Chapter 1. FcεRIβ binds allergen specific IgE expressed on Langerhans and mast cells[32]. Binding of the allergen triggers mast cell degranulation and the release of inflammatory mediators which leads to the acute phase Th2 response seen in AD. Due to the implication of MS4A2 in the inflammatory Th2 response it is a potential candidate in the pathogenesis of AD.

CMA1 produces the protein mast cell chymase 1, a chymotryptic serine proteinase which is expressed in mast cells and functions as a pro-inflammatory molecule. CMA1 is in a chromosomal region linked with hAD[43;194] and has shown associations with hAD and/or total serum IgE in three separate Japanese[132-134], one British[135] and one German population(s)[136].

Dipeptidyl-peptidase 4 (DPP4) is identical with the leukocyte surface antigen CD26 and belongs to the group of post-proline dipeptidyl aminopeptidases[191;247]. DPP4 is expressed on

the surface of resting and activated T, B and NK cells[192]. It has been implicated in the activation of regulatory T cells, T cell cycle arrest, inhibition synthesis and proliferation of T cells, and suppression of T-cell production of inflammatory cytokines[191], making DPP4 an interesting candidate for AD. Down-regulation of DPP4 has been observed in AD on a human microarray[178] and a human flow cytometery study[190]. It is suggested that reduced DPP4 expression could lead to less effective regulation of inflammation[192], delayed wound healing[193] and ultimately lesion formation in AD.

Cell cycle/apoptosis/repair/lesion formation

Cadherin-13 (CDH-13) is expressed in basal keratinocytes[248] and may modulate cell–matrix adhesion[202]. Down-regulation of CDH-13 protein was observed by immunohistochemistry in psoriasis vulgaris[249]. It is thought to act as an endogenous negative regulator of keratinocyte proliferation and its inactivation may promote the keratinocyte hyper-proliferation in psoriasis vulgaris[201]. It was found to be up-regulated in hAD using microarray experiments[178]. Because of its role in keratinocytes and increased detection on the microarray, CDH-13 represents a potential candidate gene for cAD.

Suppressor of cytokine signalling 3 (SOCS3) is a cytokine-inducible negative regulator of cytokine signalling. It is induced by a number of cytokines, of which the most relevant to AD are IL-10 and IFNγ. The chromosomal region of SOCS3 was associated with hAD in genetic linkage studies[54;115]. Furthermore, an expression study recorded significant increases in expression of SOCS3 mRNA and protein in hAD and identified haplotypes associated with human AD in a Swedish population[199]. It has been suggested that SOCS3 regulates keratinocyte proliferation and differentiation, implicating a role in skin repair and, potentially, AD[200].

Transcription Factor/miscellaneous

Peroxisome proliferator-activated receptor gamma (PPARγ) is an anti-inflammatory molecule and may be important in the development of chronic inflammation. Down-regulation of PPARγ was observed in a human microarray study of AD[178]. Reduced expression of PPARγ is associated with reduced lamellar body formation and lipid processing, and increased epidermal permeability. This could affect TEWL and barrier formation in AD[209]. However, another study using immunohistochemistry, western blotting and qPCR demonstrated that PPARγ is increased in lesional skin from AD patients[250]. The authors also demonstrated that IL-4, IL-13 and IFNγ, which are involved in the biphasic atopic immune response, directly regulate the expression of PPARγ[250]. Regulation of PPARγ therefore represents a potential novel therapeutic target AD[209].

Conclusion

Using mRNA microarray a number of candidate genes with both barrier function and immunological activity have been identified. There are similarities between this studies finding and hAD studies, suggesting that cAD is a suitable model for hAD. Using mRNA microarray and literature searches, 29 potential candidate genes were selected for further genetic research.

Chapter 4

Quantitative Real Time PCR –

Optimisation and Quality Control

4. Quantitative Real Time PCR – Optimisation and Quality Control

This work is presented in: **Wood S.H.**, Clements D.N., McEwan N.A., Nuttall T., Carter S.D. (2008), Reference genes for canine skin when using quantitative real-time PCR. *Veterinary Immunology and Immunopathology*, 126 (3-4):392-5.

4.1 Introduction

Chapter 3 identified candidate genes for further research into cAD pathogenesis by using mRNA microarray analysis. This provided useful data relating to the expression of specific genes in diseased and control skin tissues in dogs. However, as discussed previously (Chapters 1 and 3) microarrays are prone to false positive results and require confirmation by a robust quantitative method of gene (mRNA) expression. Quantitative real time PCR (qPCR) is the current "gold standard" for RNA quantification in cells or tissues and is increasingly used in veterinary research to investigate various conditions. It therefore represents the best method to confirm the mRNA microarray results in Chapter 3.

qPCR is essentially a polymerase chain reaction that is monitored in 'real time', recording the amount of PCR product generated over time instead of at one time-point, as with traditional quantification methods e.g. southern, northern blot, regular PCR. Fluorescent reporter molecules are used to monitor the amount of PCR product, the intensity is measured and a Ct value is produced. The Ct value represents the number of PCR cycles that elapse before a PCR product concentration threshold is reached. The sooner the threshold is reached the more PCR product present, allowing comparison of case and control samples.

qPCR can be absolute or relative; absolute is the most accurate method as it uses a known concentration control to calculate the number of mRNA transcripts present in the samples, whereas relative quantification compares arbitrary Ct values.

There are many pitfalls when performing qPCR experiments therefore thorough optimisation and quality control is essential to obtain accurate and meaningful results. Therefore this chapter aims to consider and describe essential quality control experiments which will allow the accurate quantification of RNA to confirm the mRNA microarray results and quantify other candidate genes potentially involved in the pathogenesis of cAD.

The comparison of qPCR data poses a number of problems because differences in tissue and mRNA storage/handling techniques[251], mRNA quality[252;253], method of reverse transcription[254], concentration of the cDNA samples and presence of PCR inhibitors can all effect the data generated. Minimisation of these differences is essential for accurate quantification of gene expression. By considering storage/handling techniques for the tissue samples and comparing subsequent mRNA quality it is possible to establish the best method for storage of tissue for RNA extraction.

Assessing the quality of the RNA is traditionally done by comparing the 28S to 18S ratio, however this method has been shown to be inconsistent as it relies on human interpretation of gel pictures[253]. To improve RNA quality assessment the RNA Integrity Number (RIN) algorithm is used[255]. RIN provides a more robust universal measure of RNA quality therefore sample comparison is standardized enabling repeatability of experiments. RIN is calculated automatically by using the entire electrophoretic trace of the RNA sample (see Figure 4.1). By applying a RIN cut-off to samples used in subsequent studies it ensures reproducibility and that only high quality RNA is used.

Figure 4.1 Examples of Electrophoretic Traces Generated to Determine RNA Integrity Number (RIN) (adapted from Schroeder *et al* (2006)[255])

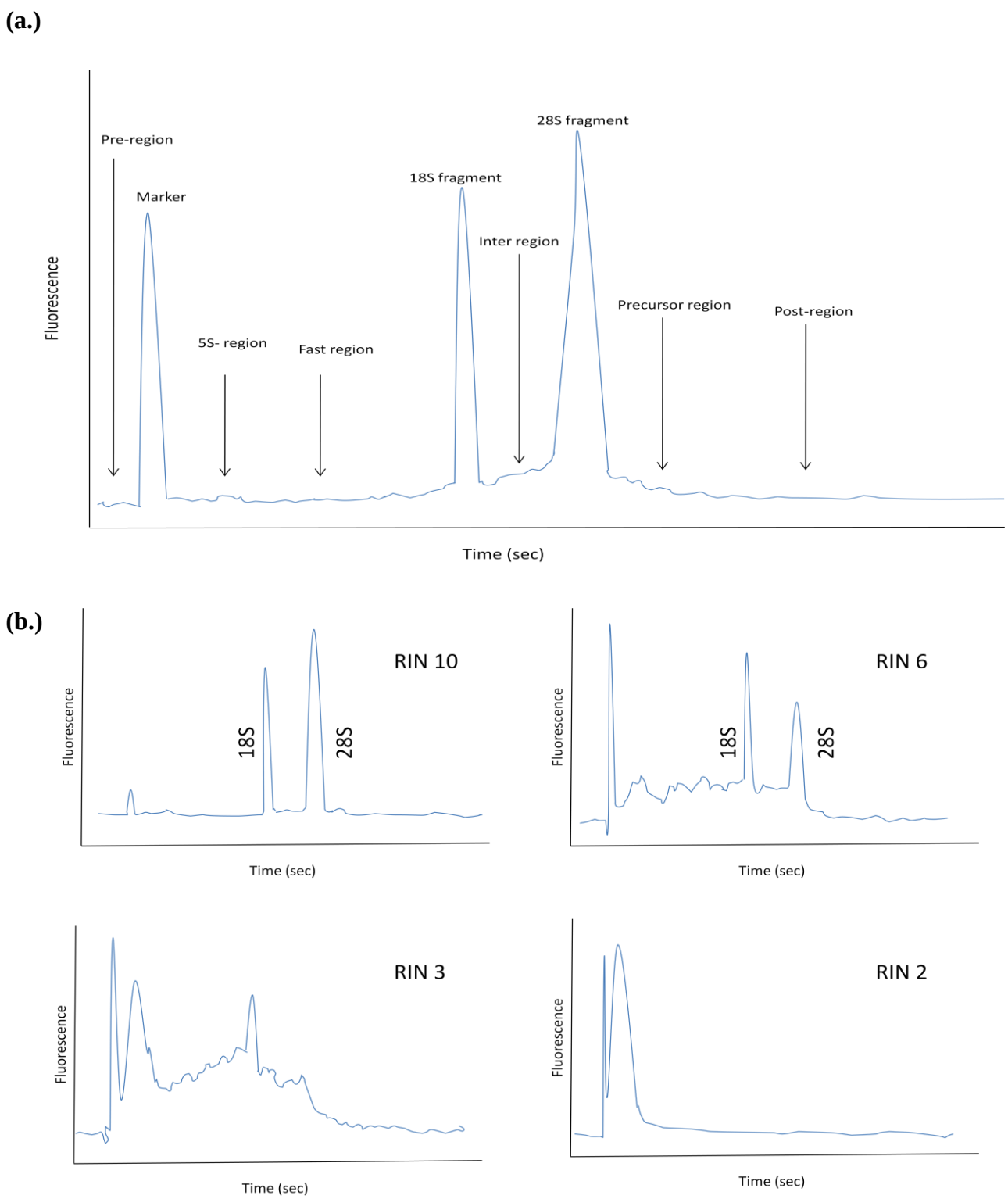


Figure 4.1: (a). Shows the areas of the electrophorectic trace that are assessed to calculate RIN, these are the 5S region, fast region, 18s fragment, inter region, 28s fragment, precursor region and the post region. (b). Shows electrophoretic traces from RNA samples ranging in RIN from 2 to 10. 10 is the best quality.

Data transformation through "normalisation" is used to correct for these differences in handling and quality. Normalisation is achieved by quantifying the expression of a constitutively expressed gene, termed a reference gene, which is unaffected by the disease process or experimental condition being investigated. Normalisation equalises variability[256], by relating the reference gene expression values to those of the gene of interest in samples analysed. Therefore, a reference gene must be stably expressed with consistent and repeatable levels of expression across all samples[257]. A reference gene that is not stably expressed could lead to errors in quantification. To improve the accuracy of results obtained from qPCR it is essential that suitable reference genes are identified and validated. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is often used as a universal reference gene for qPCR. It has, however, been demonstrated that this and other commonly used reference genes can have variable expression across tissue types and disease states[258;259]. Therefore the ideal reference gene(s) for use in a given experiment should be determined on an experiment to experiment basis, as it is unlikely there is a reference gene that is equally constitutively expressed in all tissues or cells in a given species.

Standard curves are generated in a qPCR experiment, these are necessary for accurate quantification of gene expression using qPCR. They establish the limits of detection of the assay and give the efficiency of the assay which is essential for calculating the actual transcript number when using known concentration artificial oligonucleotides for quantification of gene expression[260]. Standard curves are generated before the main experiment and occasionally are generated using a different sample type to the sample type used in the final experiment, for example using RNA extracted from keratinocyte cultures instead of RNA extracted from tissue. The reasons for this are that culture represents a freely available resource of RNA whereas experimental tissue is limited. Therefore to preserve the experimental tissue, keratinocyte RNA is used for standard curve generation. Because these

are effectively different sample types the effect on the generation of standard curves was investigated.

The first aim of this study was to establish the best tissue storage technique by comparing RIN and yield of RNA extracted. The second aim was to compare expression of candidate reference genes in canine skin to identify those suitable for this qPCR experiment. Potential reference genes for canine skin were selected from a panel of commonly used reference genes[258] and a previous microarray study that identified potential reference genes in canine tissue[257]. From this previous data IMP, CG14980, S7, HIRA, GAPDH, RPL13A and SDHA were identified as good candidates for reference genes in canine skin. The stability of the reference genes was assessed by Bestkeeper[261], GeNorm[258] and Normfinder[262] software packages. The final aim was to assess the effect of using different sample types e.g. cultured cells, clinical tissue and disease status tissues (lesional, non-lesional and control) on standard curve generation. It was considered that by performing these quality control steps the subsequent qPCR experiments would be accurate and robust.

4.2 Method

Diagnosis, sample collection, sample processing & dogs

Diagnosis, sample collection, extraction and cDNA synthesis were preformed as set out in Chapter 2 (sections 2.1, 2.2 and 2.3).

The RIN and yield comparison was performed on 57 skin samples stored by snap freezing and 22 stored in RNAlater™. The samples were a mix of lesional, non-lesional and control samples (see Appendix 1 for sample details).

The following samples were used in the reference gene experiment. Lesional skin was obtained from 3 male and 4 female atopic dogs with a mean age of 4.8 years (range 2-9),

comprising Staffordshire bull terrier, Labrador, Scottish terrier, West Highland white terrier, Boxer (2) and Neapolitan mastiff. Non-lesional skin was taken from seven atopic dogs, 4 female and 3 male, with a mean age of 1.6 years (range 11 months to 3 years). The breeds were: crossbreed, Staffordshire bull terrier, Springer spaniel, Labrador and Boxer dogs (3). The control samples were from 4 females and 3 males that had an average age of 4.8 years (range 10 months to 10 years). The breeds were: crossbreed (4), Boxer, German shepherd dog and Siberian husky.

For standard curve generation, ten-fold dilutions of cDNA derived from canine keratinocyte cultures (CELLnTEC, Bern, Switzerland), lesional skin, non-lesional skin and control skin were used. A pool of lesional, non-lesional and control cDNA skin samples were also used to generate standard curves.

Primer & Assay design

Transcript sequences and the assay design for the reference gene study were obtained from two previous studies[257;263] and are reproduced in Table 4.1. See Chapter 2 (section 2.4) for cycling conditions and assay volumes. Transcript sequences obtained from Ensembl were cross checked with NCBI canine genome transcript sequences. The assays were designed in areas that showed 100% homology between the Ensembl and NCBI sequences. To design the primers the Roche universal probe library designer was used (https://www.roche-applied-science.com/servlet/). Where possible, the primers were designed to cross an exon-exon boundary. If this was not possible, the primers were designed to be intron spanning to ensure that the primer hybridised on different exons. In these cases the intron was greater than 1000bp to maintain specificity. This approach to primer design increases the chance that the primers are transcript specific and minimises genomic DNA contamination. Finally, the specificity of the primers was checked using BLAST[164]. The primers used in the standard

curve generation were SCCA2 and Cystatin A. Table 4.1 lists the primers, probe numbers and amplicons for each gene.

The samples for standard curves were diluted serially and the assays were performed as in the Chapter 2 (section 2.4).

Table 4.1 Primers, probes and amplicons for qPCR

Gene	Forward Primer	Reverse Primer	Probe Number (Roche)	Amplicon
IMP	cgctgcctctttcaaacat	tttggcctcatcttcactgag	62	cgctgcctctttcaaacatgtcagcccagcaggtgctgctgttggaattccactcagtgaagatgaggccaaa
CG14980	gcaggaagggattctccag	gggtccagtaagaaatcttccataa	72	gcaggaagggattctccagtaagggcggagatgccaggaaatcttcaacattatggaagatttcttactggaccc
S7	agtgcagggagaagaagc	cagcagctcgtgtgacaact	89	agtgcagggagaagaagcaccggcggatgctgatgccagagaagttgtcacacgagctgctg
HIRA	aattcagaacatgctgcaatttta	tgattcatcatccataacctgttc	39	aattcagaacatgctgcaattttatattccagaggtggagggtgtagaacaggttatggatgatgaatca
GAPDH	ctggggctcacttgaaagg	caaacatggggcatcag	68	ctggggctcacttgaaaggcggggccaagagggtcatcatctctgctccttctgctgatgcccccatgtttg
RPL13A	ctgccccacaagaccaag	gggatcccatcaaacacct	64	ctgccccacaagaccaagcgaggccaggctgccctggaccgcctcaaggtgtttgatgggatccc
SDHA	ggtggcacttctacgacacc	atgtagtggatggcgtcctg	20	ggtggcacttctacgacaccgtgaagggctccgactggctgggggaccaggacgccatccactacat
SCCA2	agaattctggctggacaagg	tggtatttctaggatcttggcttg	65	agaattctggctggacaaggacacaagcaaacctgtgcagatgatgagacaatccaacgtttttaatttcacctcactggaggacttgcaagccaagatcctagaaatacca
Cystatin A	ccaagaatttgaagccgtaga	ctattatcacctaccccgcacct	83	ccaagaatttgaagccgtagagtataaaactcaagtggtggctggaataaattactacattaaggtgcgggtaggtgataatag

Comparison of RIN in snap frozen & RNAlater stored tissues

RIN and RNA yield was determined according to general methods in section 2.4. RIN and RNA yield was compared between tissues collected and stored in RNAlater and those snap frozen.

Statistical analysis

The Ct values for the reference genes were assessed for stability of expression across all samples using three different statistical algorithms; Bestkeeper[261], GeNorm[258] and Normfinder[262]. Bestkeeper uses raw Ct value for analysis but Genorm and Normfinder use the average power of Ct. To calculate the power of Ct, the E value (efficiency of the assay) is required. This is calculated using the slope of the standard curve for each gene is used:

$$E\ value = (10^{\frac{-1}{slope}} - 1) \times 2$$

$$Power\ of\ Ct = E\ value^{-Ct}$$

The consensus from these three programs was used to demonstrate the most stably expressed reference gene.

Averages of the yield and RIN were taken and the Student t-test was used to assess the difference between tissue collection and storage (for full results see Appendix 1).

The standard curves were assessed by the slope of the standard curve as established by the 7900 software and from this the percentage efficiency of the assay was calculated, as below:

$$\%\ Efficiency = (10^{\frac{-1}{slope}} - 1) \times\ 100$$

4.3 Results

Comparison of RIN

Table 4.2 RIN and yield comparison

	RNA concentration (ng/µl)	% Variation (error)	RNA Integrity Number (RIN)	% Variation (error)
RNAlater™ samples (n=22)	732.5	17.7	8.1	1.9
Snap frozen samples (n=57)	268.6	19.1	6.8	3.3

Table 4.2 shows the average RNA concentration and RIN. The percentage variation is used to represent the spread of values obtained. This is calculated using the standard error and the average value. Variation below 20% is usually acceptable in biological experiments. It is clear that the RIN has a lower percentage variation than the concentration; this is because RIN is independent of the sample concentration which is variable across samples. Using the Student T-test it was found that RNA concentration was significantly higher in RNAlater samples ($p=0.0024$) and that RIN was significantly higher in RNAlater samples ($p=0.0005$). Figure 4.2 shows the spread of RNA and RIN results for both RNA later and snap frozen samples. See Appendix 1 for full results.

Figure 4.2 RNA concentration and RNA integrity number results from Snap Frozen and RNAlater preserved skin

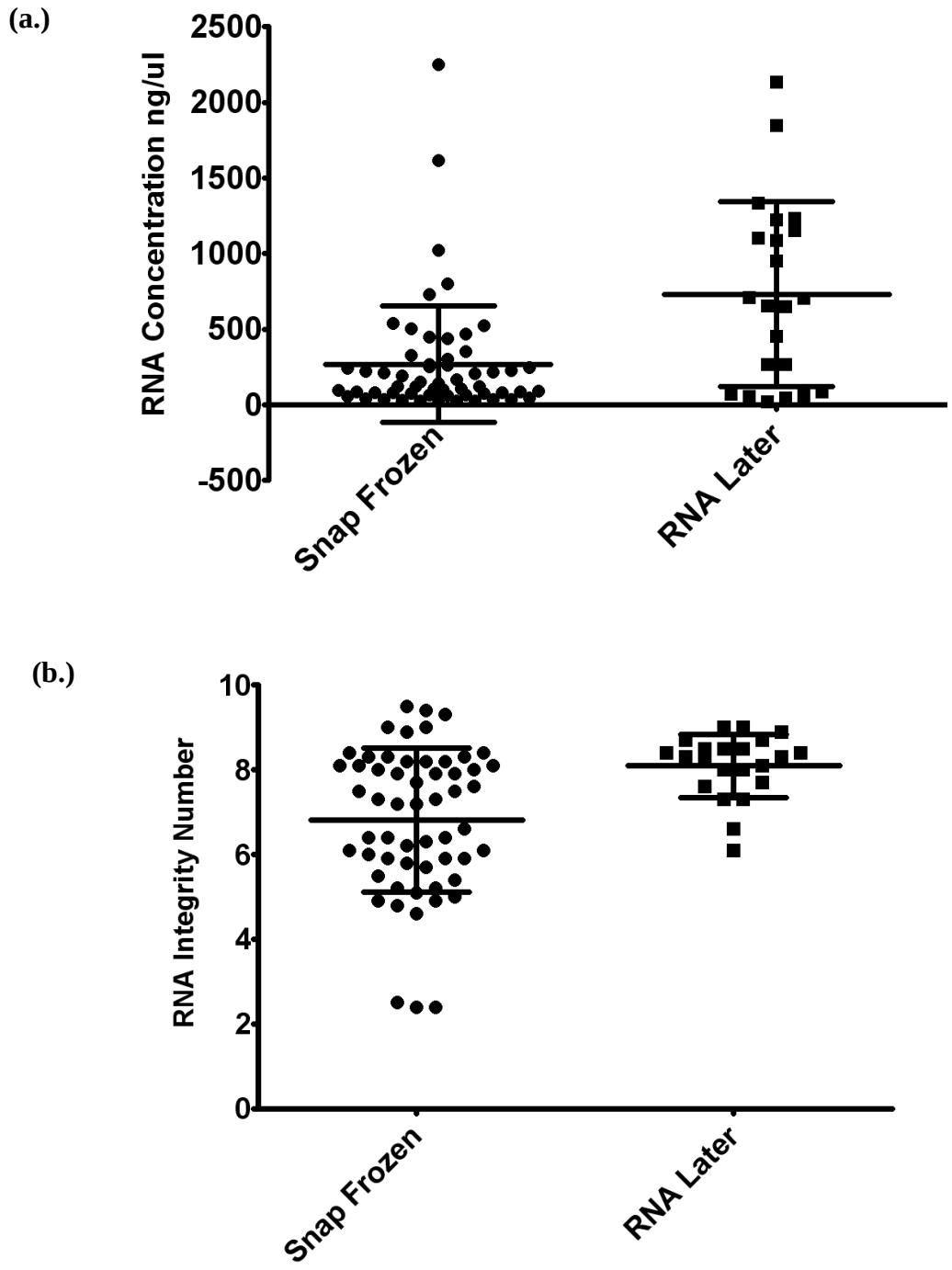


Figure 4.2: (a.) All RNA concentration values for snap frozen and RNAlater preserved skin. The mean is represented as is the standard deviation by error bars. (b.) All RNA integrity number values for snap frozen and RNAlater preserved skin. The mean is represented as is the standard deviation by error bars.

Reference gene selection

Bestkeeper

Bestkeeper (Table 4.3) determines a co-efficient of correlation, standard deviation and percentage covariance, with the user making a decision which is the best gene based on these three variables. GADPH had the best co-efficient of correlation, meaning that its expression correlated well with the other reference gene expression patterns. Nevertheless, the percentage co-variance and standard deviation were high, meaning that the stability of expression was low between the samples and replicates. SDHA and S7 had the lowest standard deviations but their correlation co-efficient was poor. Therefore, CG14980 and RPL13A offered the best stability with high co-efficient of correlation and low standard deviation and percentage co-variance.

Table 4.3 Bestkeeper results

Reference Gene	Co-efficient of correlation	Standard Deviation	% Co-Variance
CG-14980	0.992	2.008	7.20
RPL13A	0.979	1.888	10.23
GAPDH	0.99	2.756	12.12
HIRA	0.956	2.291	9.13
IMP	0.973	3.162	12.25
SDHA	0.939	1.788	7.88
S7	0.835	1.586	6.75

The co-efficient of correlation, standard deviation and percentage covariance are all outputs from Bestkeeper that represent the stability of the reference genes.

Normfinder

Normfinder (Table 4.4) returns an arbitrary stability value and standard error, based on the intra- and inter-group variations of each gene. It thereby selects the two best reference genes automatically. In this case CG14980 and RPL13A were chosen as they had the lowest stability value and standard error.

Table 4.4 Normfinder Results

Reference Gene	Stability value	Standard error
CG-14980	0.288	0.109
RPL13A	0.359	0.107
HIRA	0.534	0.116
SDHA	0.712	0.135
GAPDH	0.741	0.139
IMP	1.081	0.183
S7	1.195	0.199

Normfinder returns a stability value and standard error to represent the stability of the reference gene.

GeNorm

GeNorm (Figure 4.3) generates a graph to represent the data and calculates a stability value (M) via pair-wise comparisons. This method varies from the other two methods in that it assesses the stability of the best pair of reference genes and not individual stability. This means that the pair of genes with the most similar expression profile are selected, not necessarily the most stable gene[262]. GeNorm indicated that HIRA & RPL13A were the most stably expressed pair of reference genes, with a significant M value under 0.4[258]. CG14980 was the third most stably expressed.

Figure 4.3 GeNorm results

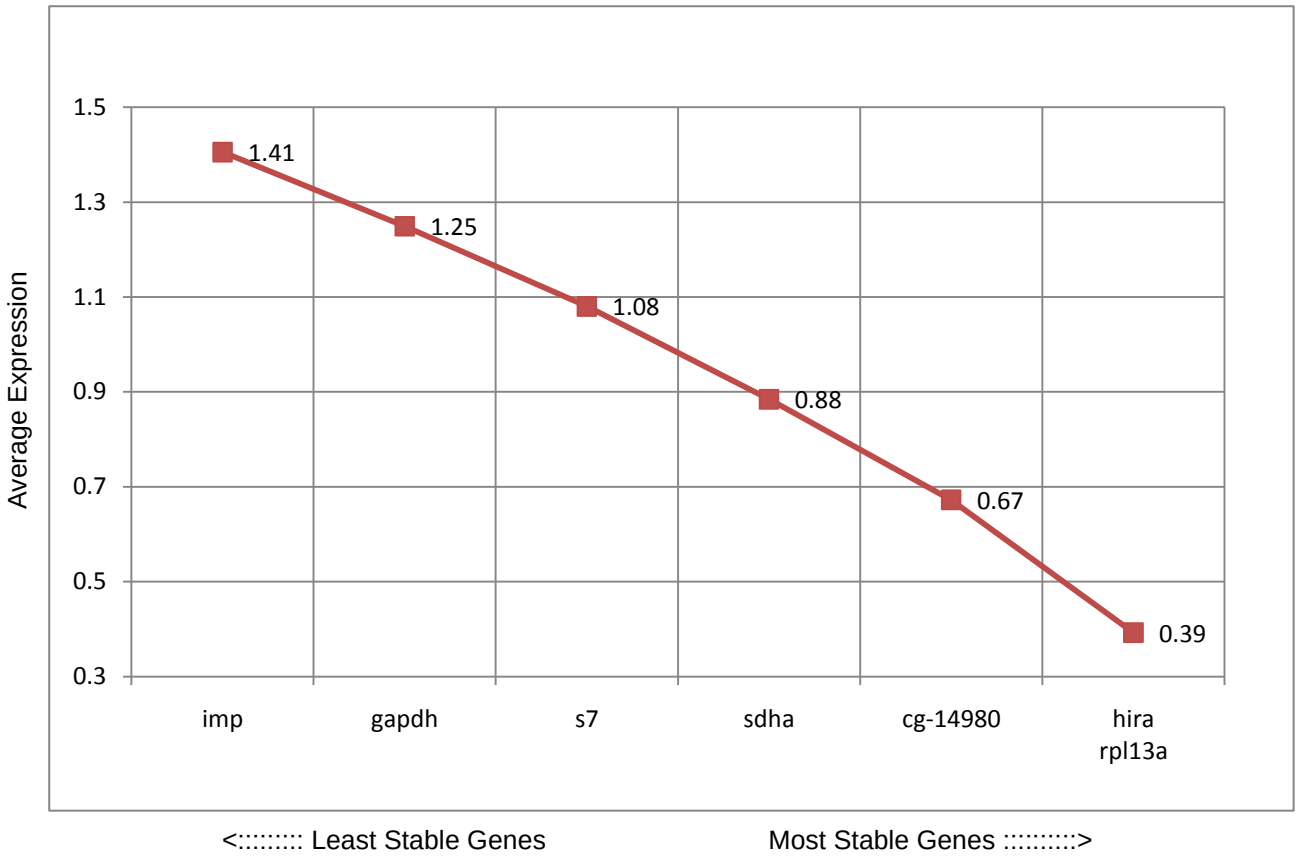


Figure 4.3 shows the output from Genorm, which plots the stability value (M) against the reference genes tested giving an order of stability.

Standard curve generation

Table 4.5 Standard curve efficiency results

Sample Type	SCCA2 efficiency (%)	Cystatin A efficiency (%)
Keratinocytes	Not expressed	100.7
Control	Not expressed	103.6
Lesional	96.1	98.7
Non-lesional	84.2	95.2
Pooled (lesion, non-lesional and control)	93.9	100.2

Table 4.5 shows that keratinocytes and healthy tissue did not express SCCA2. The efficiency of the standard curve must be within 93-107%[264;265] for accurate quantification. The SCCA2 standard curve efficiency for non-lesional skin was 84% well below acceptable limits for accurate quantification. However lesional skin had an efficiency of 96% which meant that it did fall within acceptable limits for accurate quantification. When all the samples were pooled the efficiency was 94%, which was within the acceptable limits for accurate quantification.

Cystatin A was expressed in all sample types and all fell within the efficiency parameters. However there was considerable variability with the efficiency ranging from 95% to 103%.

4.4 Discussion

It has been demonstrated that RNAlater stored skin samples are better preserved than snap frozen tissues and therefore give higher concentrations and RIN values. This suggests that in future it would be best to collect skin for qPCR in RNAlater.

Standard curve generation has shown variability in the percentage efficiency when using different sample types. Because the standard curve efficiency is required to calculate the transcript number of genes expressed it is essential that the value best represents the experimental conditions. In the case of SCCA2, keratinocytes and healthy skin did not express the gene, in this situation it would be best to use the pool of lesional, non-lesional and control skin samples because this best represents the experimental conditions and represents all sample types therefore establishing the most reliable standard curve available for this gene.

Cystatin A was expressed in all sample types; however, the percentage efficiency was variable across the sample types. Using keratinocyte cultures was the most sample efficient method for generating standard curves; however, as demonstrated by this experiment the standard curve generated by the cultures was different to the experimental sample and in the case of SCCA2 no curve could be generated. Ideally, a separate standard curve would be generated for each experimental sample type; however, this is not only time consuming and wasteful of samples and reagents, but in some cases, as demonstrated by SCCA2, no standard curve which meets the efficiency parameters can be generated. Therefore a pool of all experimental samples is the most accurate and sample efficient way to generate standard curves for further qPCR experiments.

CG14980 and RPL13A were shown by Bestkeeper and Normfinder to be the most stably expressed reference genes. HIRA was ranked 3rd and 4th by these programs

respectively, but the paired result from GeNorm found that HIRA and RPL13A were the most stably expressed. This disparity in results was due to the different methods used. Normfinder uses a model based approach that assesses each gene on the intra- and inter-group variations whereas GeNorm uses a pair-wise approach. It has been stated that the model based approach used by programs such as Normfinder is 'more precise and robust measure of gene stability than pair-wise comparisons'[262]. Also, in the case of pair-wise comparison, two genes with similar expression levels are selected, this poses a problem if candidate reference genes have different expression levels. A threshold is applied to GeNorm that avoids use of dissimilar reference genes; therefore, not only are the most stably expressed genes not selected, in some cases no genes will meet the stability threshold excluding potentially suitable reference genes. Because of the criticisms of the pair-wise approach and the consensus between Bestkeeper and Normfinder, it was determined that RPL13A and CG14980 were the most suitable reference genes from the panel of genes analysed for use with healthy and atopic canine skin. This study validates these reference genes for use in canine skin for future qPCR studies.

The results also determined that GAPDH was not stably expressed in canine skin. It is therefore probable that reliance on this reference gene may have influenced the results of previous qPCR studies in canine skin. Clearly, it would be pertinent to determine how the use of CG14980 and RPL13A, compared to GADPH, alters the results obtained from real time experiments.

There is a possibility that these results have wider application, as it has been suggested that reference genes are relatively species independent[266]. If this is the case then these reference genes may be suitable for skin research in other species including humans. However, it has been demonstrated that the feline and canine reference genes differ in

stability[267], although this analysis was based on GeNorm pair-wise comparisons which as stated earlier may not always select the most stable reference genes.

It is important to state that these reference genes are not necessarily optimal for use with other tissue types, even those derived from skin such as keratinocytes. Reference genes should be tested prior to qPCR studies and the most stable for the specific tissue type chosen. In addition, the use of one reference gene is discouraged as it has previously been demonstrated that the use of two to four reference genes improves accuracy of quantification[258;266;268].

In conclusion, it is best to collect and store skin samples in RNAlater (compared to snap frozen tissues); also the use of pooled samples for standard curve generation is recommended. And finally in cAD when performing qPCR on skin samples the use of CG14980 and RPL13A as reference genes is advised.

Chapter 5

Quantitative Real Time PCR - Analysis of Gene Expression in Canine Atopic Dermatitis and Correlations with Clinical Severity Scores

5. Quantitative Real Time PCR - Analysis of Gene Expression in Canine Atopic Dermatitis and Correlations with Clinical Severity Scores

This work is presented in: **Wood S.H.**, Clements D.N., Ollier W.E., Nuttall T., McEwan N.A., Carter S.D. (2009) Gene expression in canine atopic dermatitis and correlation with clinical severity scores. *Journal of Dermatological Science*, 55(1):27-33

5.1 Introduction

Chapter 4 detailed the quality control experiments undertaken to establish a robust methodology for quantification of RNA in canine skin samples. Using this methodology, this chapter will detail the qPCR experiments carried out to validate gene expression data from the microarray study (Chapter 3) and to investigate changes in gene expression in candidate genes identified from literature searches on the pathogenesis of canine and human AD (Chapter 3).

Validation of the mRNA microarray results is necessary because mRNA microarrays are prone to false positive results[158]. Moreover, mRNA microarrays are relatively insensitive to small changes in expression and, therefore, relevant genes could be missed.

Figure 5.1 shows the selection of genes for this study using Table 3.4. Twelve genes with relevant epidermal or immune functions (Cystatin A[235;236], CARD4[75] P-selectin[269;270], PKP2[176], PPARγ[250;271], SGPL1[171], TNF-α[43;272], Cadherin-13[178], CMA1[132;136], DPP4[190], SPINK5[126;127], SAA-1[273]) were selected as potential candidate genes for cAD following a literature search. Eight further genes (ARTS-1, Cullin-4A, INPPL1, S100A8, POSTN, SCCA2, STAT2, TIMP-1) were selected using data from the canine mRNA expression microarray[274] (Chapter 3) that indicated they were dysregulated in cAD.

Figure 5.1 Gene selection for q-PCR quantification

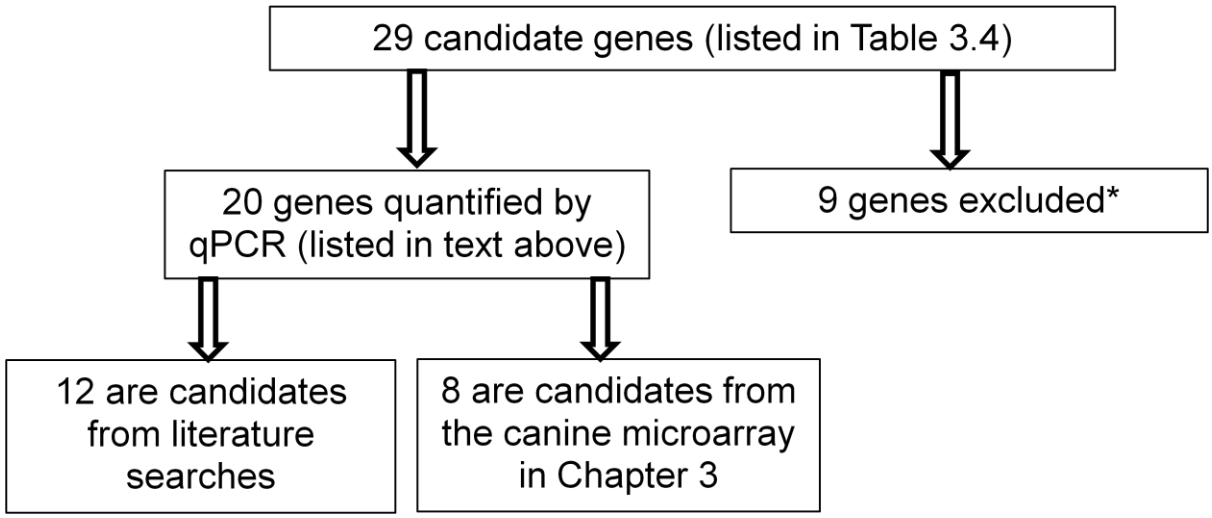


Figure 5.1 From the 29 genes listed in Table 3.4, 20 were selected for quantification by qPCR, 12 of which were from literature searches and 8 were dysregulated on the cAD microarray. Nine genes were excluded because primers or assays within acceptable limits could not be designed. The excluded genes were: GOLGA4, MUC2, TSLP, SPRR1B, IL1RL1, CTLA4, SYND1, SOCS3 and Filaggrin.

Optimisation and quality control of data are essential in qPCR experiments and therefore, CG14980 and RPL13A[275] were chosen as reference genes and pooled samples were used to generate standard curves and give the results more accuracy and validity.

It was hoped that quantifying gene expression in atopic skin would increase understanding of the complex gene interactions in this condition and identify causative genes involved in disease pathogenesis.

Canine Atopic Dermatitis Severity Index (CADESI-03) scores were recorded for each atopic dog. This is a well validated assessment of clinical lesions (erythema, excoriation, lichenification and self-induced alopecia) at 62 anatomical sites from 0 (normal) to 5 (most severe) yielding a score of 0-1240 (similar to the human SCORAD index)[153]. This assessment provides a score summarising the severity of the clinical signs in the affected dogs. In addition, the number of positive reactions to an intra-dermal allergen test (IDT) using 54 environmental allergens (Greer Laboratories Inc., Lenoir, NC, USA) (similar to prick tests in

human patients) were recorded, as a measure of allergen specific IgE. In contrast to humans, total IgE levels in dogs do not correlate with atopic status or clinical signs, and the specificity and sensitivity of IgE specific serology is variable[33;276].

The first aim of this study was to quantify the expression of the selected candidate genes in lesional atopic, non-lesional atopic and healthy canine skin. The second aim was to correlate gene expression in atopic skin with the two clinical measures: the CADESI-03 score and the number of positive reactions on IDT. This dual approach was designed to help identify gene changes that were relevant to the clinical cause and outcome of AD. Furthermore, it was hoped that these experiments would help to determine whether cAD and hAD share a similar genetic background and help identify novel target genes for further study of pathogenesis and intervention.

5.2 Method

Diagnosis, sample collection, sample processing & dogs

Diagnosis, sample collection, extraction and cDNA synthesis was preformed as described in Chapter 2.

Lesional Skin was obtained from 17 affected dogs (8 male and 9 female) with a mean age of 3.9 years (range 11 months to 9 years). The breeds were: Boxer (5), Staffordshire bull terrier (2), Labrador (2), crossbreed (2), Springer spaniel, Rhodesian ridgeback, Scottish terrier, bulldog, West Highland white terrier and Neapolitan mastiff.

Fourteen of the non-lesional samples were taken from the same dogs as the lesional samples. Three samples were taken from different dogs and the non-lesional age range was 11 months to 8 years with an average age of 3.4 years. Nine were male and 8 were female and the breeds were as follows: Boxer (5), crossbreed (3), Staffordshire bull terrier (2),

Springer spaniel, Labrador (2), Rhodesian ridgeback, bulldog, West Highland white terrier and Neapolitan mastiff.

The control dogs had a mean age of 4.8 years (range 10 months to 10 years) and comprised 10 females and 7 males. The breeds were: crossbreed (14), Boxer, German shepherd dog, Siberian husky.

Primer & Assay design

PCR cycling conditions and assay volumes are described earlier in Chapter 2.4. Transcript sequences obtained from Ensembl[159;160] were cross checked with NCBI canine genome transcript sequences. The assays were designed for sequence areas that showed 100% homology between the Ensembl and NCBI sequences. To design the primers the Roche universal probe library designer was used[277]. Wherever possible, the primers were designed to cross an exon-exon boundary. If this was not possible, the primers were designed to be intron spanning to ensure that the primer hybridised on different exons. In these cases the intron was greater than 1000bp to maintain specificity. This approach to primer design increased the chance that the primers were transcript specific and minimised genomic DNA contamination. Finally, the specificity of the primers was checked using BLAST[164]. Table 5.1 lists the primers, probe numbers and amplicons for each gene (see Appendix 2 for R^2, slope, percentage efficiency and E value for each assay. These values were used in the statistical analysis and were calculated in advance to ensure the assay was within acceptable limits of detection and the efficiency range of 97-103%).

Synthesised oligos of the amplified region (supplied by Eurogentec) for each gene were used to calibrate and allow absolute quantification of transcript number in all samples as previously described[260]. A full explanation of data transformation is presented in this section under the heading “statistical analysis” (p.94).

Table 5.1 Primers, probes and amplicons for qPCR

Gene	Forward Primer	Reverse Primer	Probe Number (Roche)	Amplicon
SCCA2	agaattctggctggacaagg	tggtatttctaggatcttggcttg	65	agaattctggctggacaaggacacaagcaaacctgtgcagatgatgagacaatccaacgtttttaatttcacctcactggaggacttgcaagccaagatcctagaaatacca
Cystatin A	ccaagaatttgaagccgtaga	ctattatcacctaccccgcacct	83	ccaagaatttgaagccgtagagtataaaactcaagtggtggctggaataaattactacattaaggtgcgggtaggtgataatag
Card 4	tcgtcctgcaccactcc	tgatctggtttacgctgagtct	85	tcgtcctgcaccacttccgcaagcggcttgccctcgacctggacaacaacaatctcaacgactacggcgtgagggagctgcagccctgcttcagccgcctcaccgtcctcagactcagcgtaaaccagatca
P-selectin	ctgcaccaatctgcaaagc	atgagggctggacactgaac	17	ctgcaccaatctgcaaagcaggcatagtgtcagctcctacttcaaaggttcagtgtccagcccctcat
PKP2	aagcatctttgggagctctg	ggccattttccttctggac	4	aagcatctttgggagctctgcagaatctcactgcaggaagtggaccaatgccgacatcagtagctcagacagttgtccagaaggaaaatggcc
PPARɣ	caggaaagacaacagacaaatca	ggggtgatgtgtttgaacttg	7	caggaaagacaacagacaaatcaccatttgttatctatgacatgaattccttaatgatgggagaagataaaatcaagttcaaacacatcacccc
SGPL1	cgggttccactgaacaaaat	ggagatggctctcctcattg	10	cgggttccactgaacaaaatgatggaggtggatgttcgggcaatgaggagagccatctcc
TNF-α	atggcctccaactaatcagc	cttggggttcgagaagatga	61	atggcctccaactaatcagccctcttgcccagacagtcaaatcatcttctcgaaccccaag
Cadherin 13	gccctcttcctggcaatc	tccagagttatcagcaaagttcc	59	gccctcttcctggcaatcgacagtggcaaccctcccgctaccggtaccggaactttgctgataactctgga
CMA1	aggcggaacttcgtactgac	cccagggtgaccattatga	4	aggcggaacttcgtactgacagctgctcactgtgcaggaaggttcataatggtcaccctggg
DPP4	agacgcaaagtactatcaactgagat	gctgctcctatgcagggtat	81	agacgcaaagtactatcaactgagatgttcaggccctggtctgcccctctataccctgcataggagcagc

Table 5.1 continued

POSTN	gggaagaacgaatcattacagg	ttgcaacaatttcttcagagtttc	77	gggaagaacgaatcattacaggtcctgaaataaaatatactaggatttctactggtggtggagaaacagaagaaactctgaagaaattgttgcaa
SPINK5	gaaagaggaggacaacttgagaa	gaattcgtggcactgatcct	2	gaaagaggaggacaacttgagaaacacaggagaaaagagtaatgaaaaccaggatcagtgccacgaattc
SAA-1	ttgtgctccctggtcctg	gagtaggctctccacatgtctct	66	ttgtgctccctggtcctgggtgtcagcagccagagatggttgacattcctcaaggaagcgggtcaagggactagagacatgtggagagcctactc
ARTS-1	cctcatctgtccacgtctga	tgaagtggaaaatcagttcaagg	66	tgaagtggaaaatcagttcaaggcctttctcatcaggctgctgagggacctcattgataatcagacgtggacagatgagg
INPPL1	tctcgaagctcttcttgtactcc	cgcaccaagttcttcattgag	39	cgcaccaagttcttcattgagttctactccacctgcctggaggagtacaagaagagcttcgaga
Cullin4A	ccttggagagttccatgtcc	tctatgttgtcaaaactaaagcatga	11	tctatgttgtcaaaactaaagcatgaatgcggcgctgctttcaccagcaagctggaaggcatgttcaaggacatggaactctccaagg
TIMP-1	gtggggcacaggtacagg	cccagagagactcaccagaga	3	cccagagagactcaccagagaacccaccatggcacccttgcgcccctggcctcctgcatcctgctgttgctgtggctgaccgcccccagcagggcctgtacctgtgccccac
S100A8	caatgagggagttatggcact	aaacctggtggggcagat	154	aaacctggtggggcagatccttgggcaccatgctgacggaactggagagtgccataaactccctcattg
STAT2	tctccagctccaaggactct	aggctcattgtggtctctaacag	36	aggctcattgtggtctctaacagacaggtggatgagctgcaacaaccgctggagcttaagccggagccagaagcagagtccttggagctggaga

Statistical analysis

Normalised transcript numbers obtained from calibration with synthesised oligos and reference genes were used for the statistical analysis to compensate for the differences in concentration of the samples and accurately quantify expression. To calculate normalized transcript numbers the following method was used:

$$Volume\ of\ synthesised\ oligo\ \times Molar\ concentration\ \times Avagadro's\ Number = Calibrator\ template\ number$$

The Ct difference between the synthesized oligo and experimental sample was calculated:

$$Ct\ synthesized\ oligo - Ct\ experimental\ sample = Ct\ difference$$

E power was derived from the E value of the assay (see Chapter 4 Section 4.2, statistical analysis for calculation of E value); using Ct difference a power calculation is performed:

$$E\ power = Evalue^{Ct\ difference}$$

The transcript number was calculated using the E power and calibrator template number:

$$E\ power\ \times calibrator\ template\ number = transcript\ number$$

To take into account the differences in RNA concentration the samples were normalized. The geometric means of the reference gene transcript numbers were used to do this, for each experimental sample in turn:

$$\sqrt{RPL13A\ x\ CG14980} = geometric\ mean\ of\ reference\ genes$$

$$\frac{Reference\ gene\ geometric\ mean}{Transcript\ number\ of\ experimental\ sample} = Normalised\ transcript\ number$$

Normalised transcript number was used in all subsequent statistical calculations (see Appendix 2 for E values, percentage efficiency, slope and R^2 values for each gene assayed).

The data were considered to be not normally distributed and therefore non-parametric tests were employed. Wilcoxon ranked sign tests (systat 12) were used to compare matched lesional and non-lesional samples. Mann Whitney U tests were used to compare data from atopic and control samples. The level of statistical significance was set at $p<0.05$

The correlation analysis was carried out using the Pearson R test (Excel) which gives a correlation coefficient and a p value (<0.05).

Fold change was used to compare microarray and qPCR results:

$$\frac{Mean\ transcript\ number\ (disease)}{Mean\ transcript\ number(control)} = Fold\ Change\ (y)$$

The fold change was transformed to represent the true fold change this process is different if gene is up-regulated or down-regulated.

Up-regulated gene correction to obtain true fold change:

$$y - 1 = true\ fold\ change$$

Down-regulated gene correction to obtain true fold change:

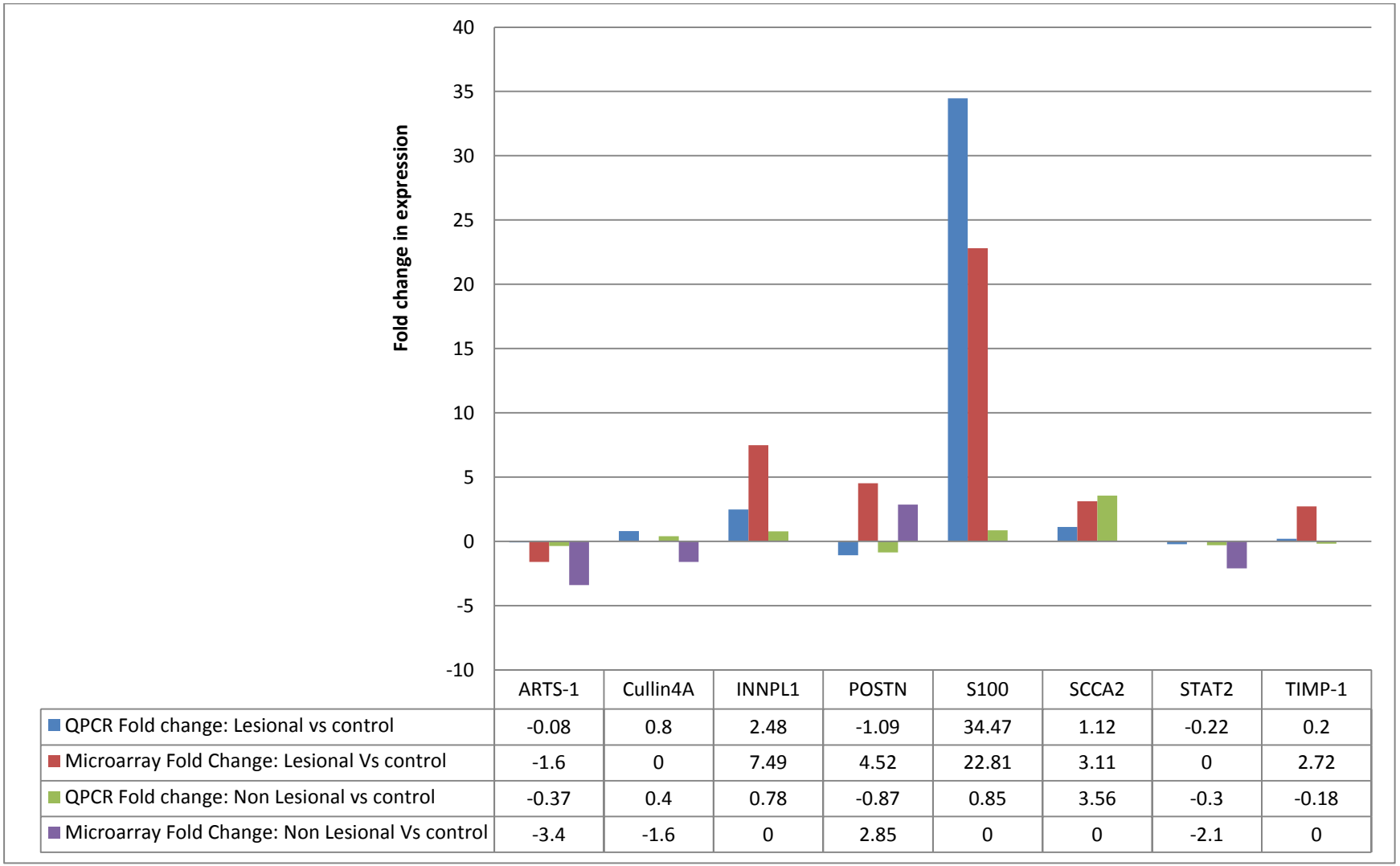
$$(-1 \div y) + 1 = true\ fold\ change$$

5.3 Results

Validation of microarray results

In the majority of cases, the microarray data overestimated the fold change, apart from S100A8 where it was underestimated (Figure 5.1). INPPL1, S100A8 and ARTS-1 showed the same trend of expression in qPCR and microarray, whereas Cullin4A and POSTN displayed opposing trends. qPCR, but not microarray, revealed a fold change in SCCA2 (non-lesional), S100 (non-lesional), INPPL1 (non-lesional), TIMP-1 (non-lesional), Cullin 4A (lesional) and STAT2 (lesional). Conversely, microarray, but not qPCR, revealed fold changes in TIMP-1 (lesional).

Figure 5.2 Comparison of microarray fold change to qPCR fold change



	ARTS-1	Cullin4A	INNPL1	POSTN	S100	SCCA2	STAT2	TIMP-1
QPCR Fold change: Lesional vs control	-0.08	0.8	2.48	-1.09	34.47	1.12	-0.22	0.2
Microarray Fold Change: Lesional Vs control	-1.6	0	7.49	4.52	22.81	3.11	0	2.72
QPCR Fold change: Non Lesional vs control	-0.37	0.4	0.78	-0.87	0.85	3.56	-0.3	-0.18
Microarray Fold Change: Non Lesional Vs control	-3.4	-1.6	0	2.85	0	0	-2.1	0

Quantification of candidate genes

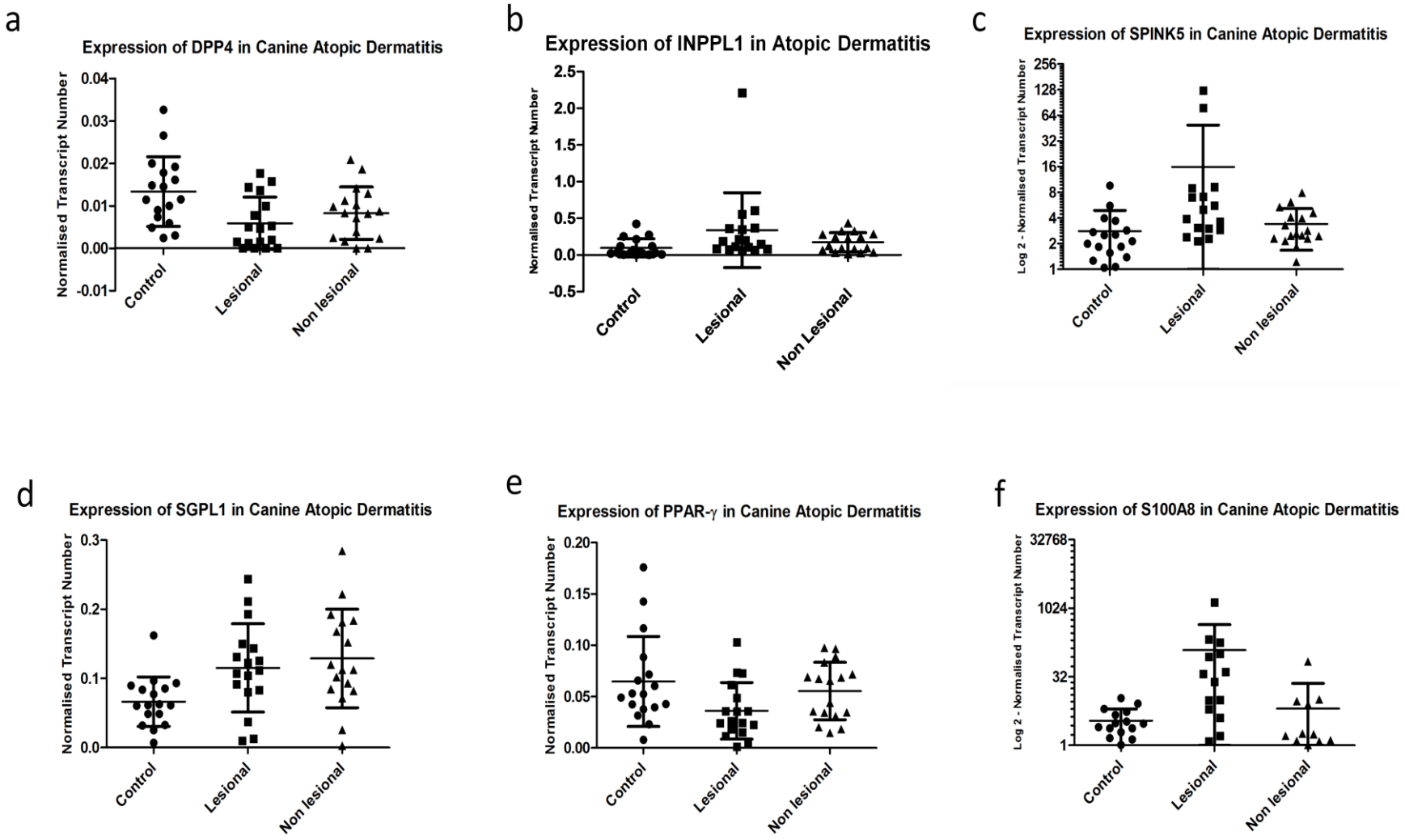
Table 5.2 shows the fold change in gene-specific mRNA expression between lesional, non-lesional and control skin. Figure 5.2 demonstrates these changes in expression for all the statistically significant genes for each individual sampled. Statistically significant changes in fold change of gene expression were seen between lesional atopic and control skin for SPINK5, INPPL1, DPP4, SGPL1, PPARγ, S100A8, PKP2, POSTN and Cullin-4A. Non-lesional skin also showed statistically significant changes in expression compared to control skin for DPP4, INPPL1, PPARγ and S100A8, though to a lesser extent than in lesional skin. Only PPARγ and S100 showed a statistically significant difference in expression in lesional and non-lesional skin. TNF-α was down-regulated in non-lesional skin compared to controls and lesional skin, although this was only statistically significant when compared to controls. There were large changes in expression of SAA-1 and SCCA2 in lesional and non-lesional skin respectively compared to control skin but these were not statistically significant due to the large variation between individual cases (e.g. percentage variation SCCA2 = 31-84%; SAA-1 = 19-46%).

Table 5.2 Gene expression in lesional, non-lesional and control skin

Gene	Lesional vs. Control skin ≠	P value	Non-lesional vs. Control skin ≠	P value	Non-lesional vs. Lesional skin ≠	P value
DPP4	-1.11*	0.002	-0.53*	0.03	-0.38	0.12
INPPL1	2.48*	0.002	0.78*	0.01	0.95	0.21
SPINK5	4.68*	0.004	0.22	0.07	3.66	0.10
SGPL1	0.74*	0.004	0.94*	0.001	-0.12	0.11
PPARγ	-0.79*	0.01	-0.17	0.39	-0.53*	0.03
S100A8	34.47*	0.02	0.85*	0.05	18.15*	0.02
PKP2	0.38*	0.03	0.52*	0.001	-0.10	0.95
POSTN	-1.09*	0.03	-0.87	0.08	-0.12	0.45
Cullin4A	0.80*	0.03	0.40	0.28	0.28	0.45
Cadherin-13	-0.51	0.06	-0.25	0.12	-0.21	0.33
TNF-α	-0.71	0.12	-0.92*	0.04	0.12	0.16
SAA-1	2.39	0.17	0.13	0.11	2.00	0.15
P-selectin	-0.22	0.18	-0.32	0.08	0.08	0.29
TIMP-1	0.20	0.23	-0.18	0.83	0.42*	0.05
SCCA2	1.12	0.27	3.56	0.44	-1.14	0.50
STAT2	-0.22	0.32	-0.30	0.19	0.06	0.21
Cystatin-A	0.07	0.33	0.37	0.43	-0.28	0.16
CMA1	0.15	0.38	0.56	0.17	-0.36	0.29
CARD4	-0.13	0.39	-0.04	0.27	-0.09	0.54
ARTS-1	-0.08	0.49	-0.37	0.11	0.27	0.21

The ≠ denotes that the figures given are true fold change. Genes highlighted in yellow show genes that have statistically altered gene expression, the * denotes which change is statistically significant (<0.05).

Figure 5.3 Scatter Plots of Individual Gene Expression for Control, Non-lesional and Lesional Skin in Atopic Dermatitis



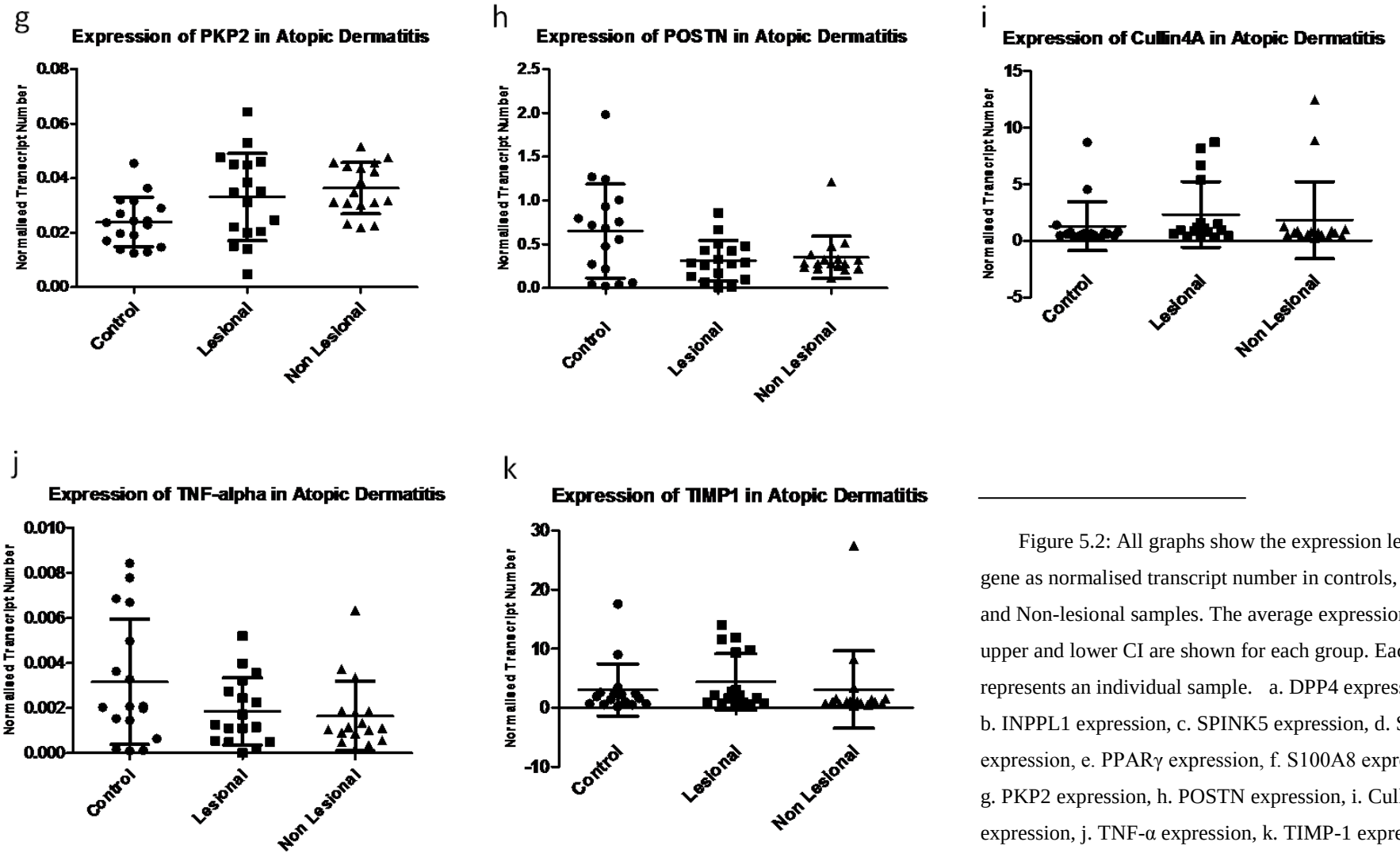


Figure 5.2: All graphs show the expression level of a gene as normalised transcript number in controls, lesional, and Non-lesional samples. The average expression and the upper and lower CI are shown for each group. Each point represents an individual sample. a. DPP4 expression, b. INPPL1 expression, c. SPINK5 expression, d. SGPL1 expression, e. PPARγ expression, f. S100A8 expression, g. PKP2 expression, h. POSTN expression, i. Cullin4A expression, j. TNF-α expression, k. TIMP-1 expression

Correlation with IDT

The majority of dogs included in this study had at least 5 positive reactions on the IDT. Figure 5.3a shows that CMA1 expression correlated with IDT data. CMA1 is a pro-inflammatory molecule and increased expression would have been expected in AD; this, however, was not the case (Table 5.2). SAA-1, also a pro-inflammatory molecule, shows a correlation with IDT (Figure 5.3c). This too did not show a statistically significant change in expression in cAD samples (Table 5.2). S100A8 is a pro-inflammatory molecule that showed a strong correlation with IDT (Figure 5.3b). The range of IDT score limits the efficacy of the results. Nevertheless S100A8 expression was shown to be significantly increased in cAD (Table 5.2), further suggesting its involvement in the pathogenesis of AD. Finally, SPINK5 expression showed a positive correlation with IDT (Figure 5.3d), although this was weaker than previous correlations with pro-inflammatory molecules.

Figure 5.4 Correlation between the total number of positive reactions on an IDT with 54 environmental allergens and normalised transcript number in non-lesional atopic skin.

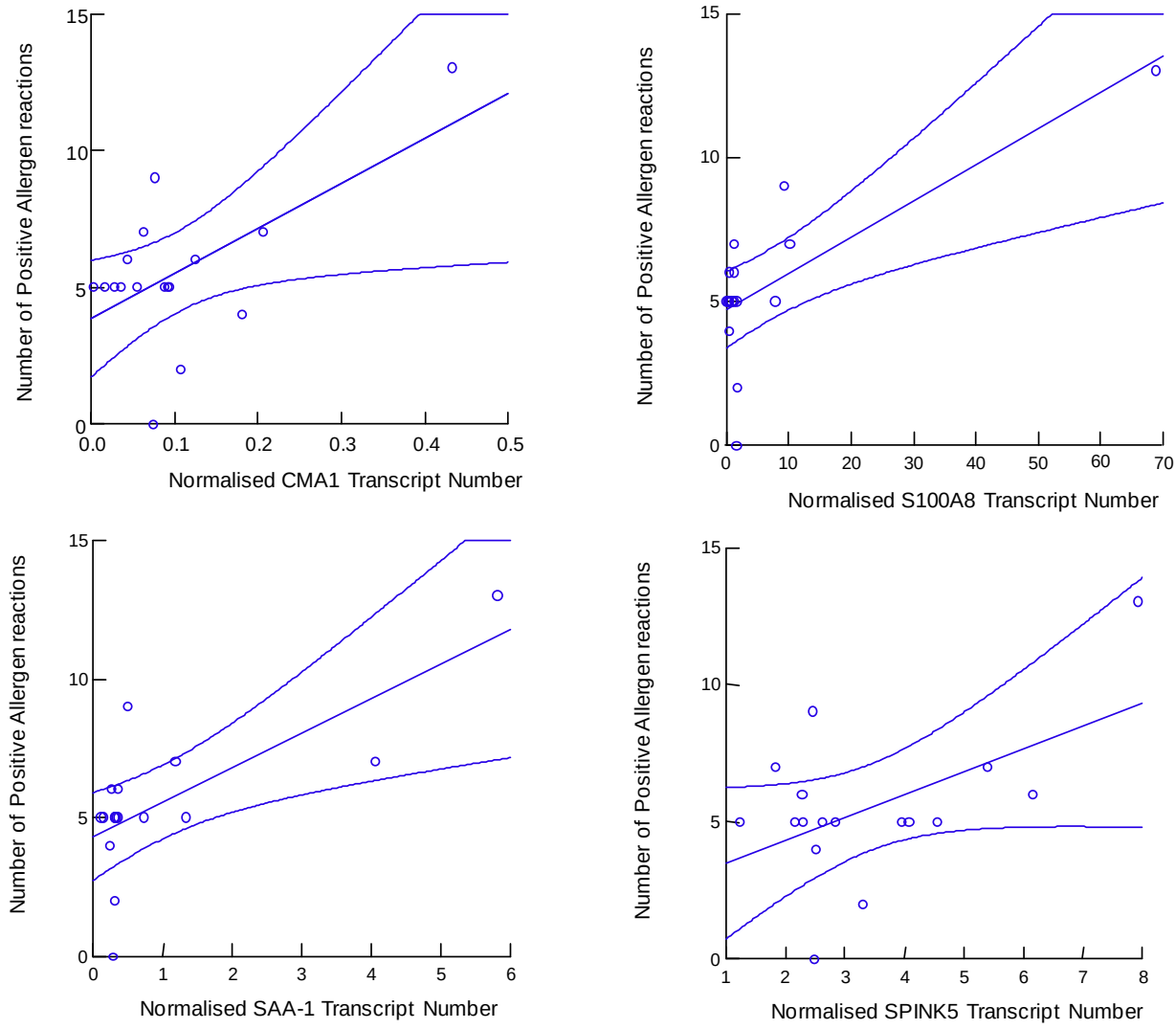


Figure 5.3: The lines on the graph indicate the direction of the relationship and the 95% confidence intervals. (a). CMA1 normalised transcript number in non-lesional atopic skin (Pearson's R value 0.607; =0.01). (b). S100A8 normalised transcript number in non-lesional atopic skin (Pearson's R value 0.751; p=0.001). (c). SAA-1 normalised transcript number in non-lesional atopic skin (Pearson's R value 0.706; p=0.002). (d). SPINK-5 normalised transcript number in non-lesional atopic skin (Pearson's R value 0.53; p=0.021).

Correlation with CADESI

Figure 5.5 Correlation between CADESI-03 and Normalised Transcript Number in non-lesional atopic skin.

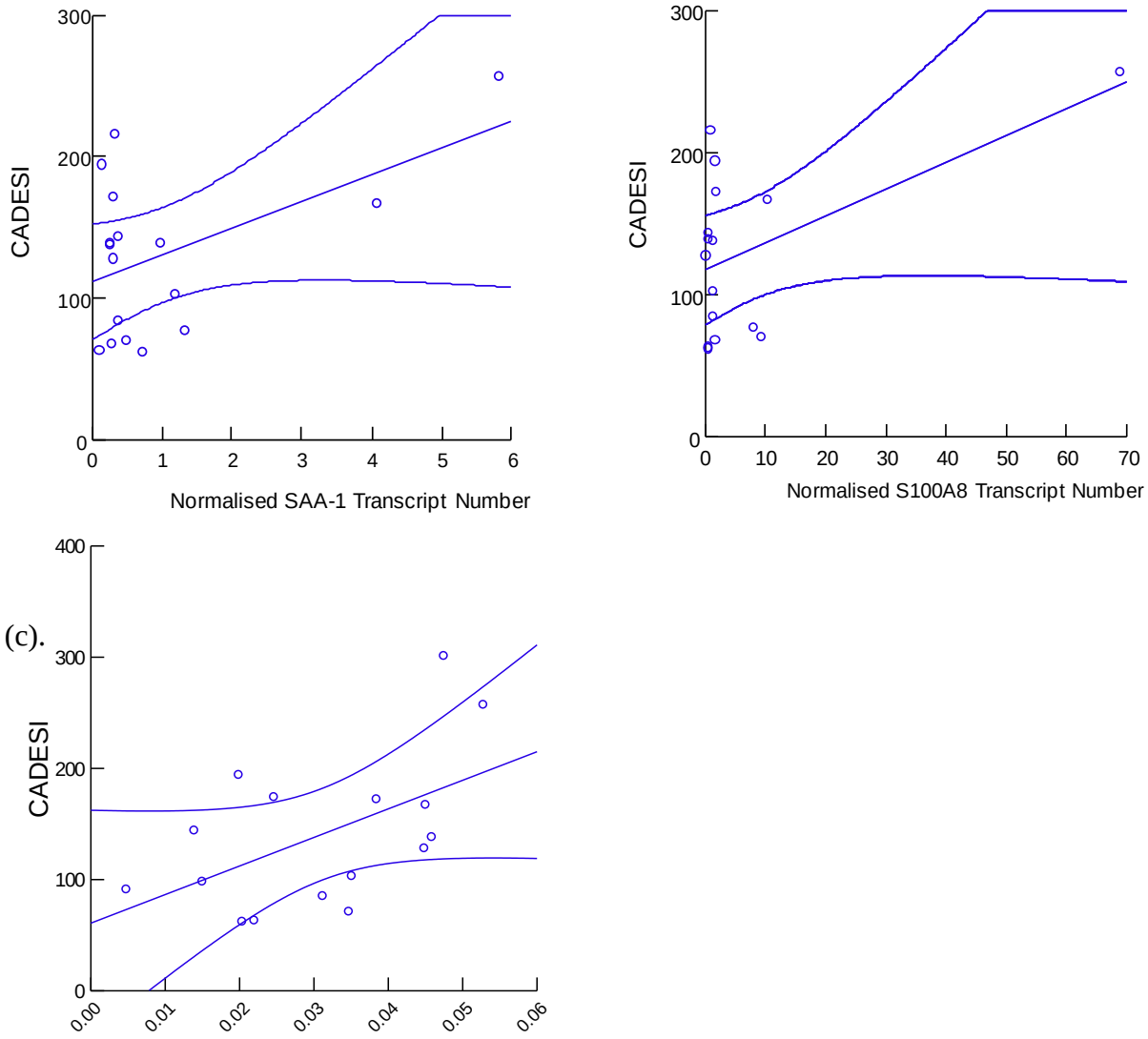


Figure 5.4: The lines on the graph indicate the direction of the relationship and the 95% confidence intervals. (a). SAA-1 Normalised Transcript Number in non-lesional atopic skin (Pearson's R value 0.505 ; p=0.039). (b). S100A8 Normalised Transcript Number in non-lesional atopic skin (Pearson's R value 0.54 ; p=0.031). (c). PKP2 Normalised Transcript Number in lesional atopic skin (Pearson's R value 0.532; p=0.034).

SAA-1 showed a correlation with CADESI-03 (Figure 5.4a) as well as IDT, although the 'r-value' was lower than with IDT (possibly because CADESI-03 assesses a number of variables whereas the IDT is positive or negative). S100A8 also showed a correlation with CADESI-03 (Figure 5.4b), again weaker than its correlation with IDT. PKP2 showed a correlation with CADESI-03 (Figure 5.4c). PKP2 expression is also increased in AD (Table 5.2).

5.4 Discussion

Assessments of gene expression in health and disease have proven to be useful in studying the complex pathogenesis of many multi-factorial diseases. Relating such changes to measures of clinical progress/disease are less common but help to better understand the underlying disease mechanisms.

Data from previous microarray studies and from the published literature (Chapter 3) were used to select candidate genes implicated in the pathogenesis of cAD. The data showed that 11 out of the 20 quantified genes demonstrated statistically significant altered mRNA expression between atopic and healthy skin. The functions of these genes vary, but include both immunological and skin barrier functions. Seven of the 11 genes have shown previous associations or altered expression in hAD. The remaining four genes were shown to be dysregulated in cAD on the canine microarray study in Chapter 3[274]. Three genes positively correlated with CADESI-03 scores and four with IDT results. The functions of these genes vary, but include both immunological and skin barrier functions.

When comparing qPCR data with the microarray data, mixed results were generated; TIMP-1, POSTN and Cullin4A probably represent false positives on the microarray, because the qPCR results contradicted the microarray findings. ARTS-1, INPPL1 and S100A8, however, showed a similar trend in expression between qPCR and microarray, corroborating

evidence of changed expression of these genes in cAD. SCCA2 and STAT2 were not statistically significantly dysregulated in the qPCR experiment therefore no conclusions could be made with regards to their role in cAD. The detection of potential false positive findings demonstrated the importance of qPCR validation of microarray results.

S100A8 levels correlated with CADESI-03, suggesting that S100A8 expression in dogs correlates with disease severity. Moreover, expression was markedly increased in lesional atopic skin compared to both non-lesional atopic (18.15 fold change) and healthy skin (34.5 fold change). Expression in non-lesional atopic skin correlated with the number of positive reactions on an IDT, although the clinical relevance of this is unclear. S100A8 is a pro-inflammatory molecule released by phagocytes in inflammatory conditions[187] which 'tracks' the severity of inflammatory diseases in humans[210]. S100A8 has also been linked to leukocyte recruitment, suggesting it is a key gene in the inflammatory response[187]. Due to the involvement of S100A8 in the immune response and its correlation with severity of symptoms these results suggest that S100A8 could be involved in the development of Th2 mediated acute inflammation and Th1 mediated chronic inflammation in cAD as described in the biphasic response model[26]. The elevated expression of S100A8 in non-lesional skin could be due to infiltration of inflammatory cells from lesional sites, evidence of the early stages of lesion formation or possibly be an indication of cutaneous hyper-reactivity in non-lesional skin as well as lesional skin in dogs with cAD.

INPPL1 mRNA expression was also up-regulated in both lesional (2.48 fold change) and non-lesional (0.78 fold change) atopic skin compared to healthy skin, suggesting that it may be a good candidate for further research. In humans, INPPL1 has roles in calcium transport, B cell activation, FcεRIβ mediated mast cell and Langerhans cell activation, regulation of IgE mediated IL-6 production[183] and T cell development[278]. Dysregulation of INPPL1 could

therefore affect allergen specific IgE responses, allergen presentation and uptake, T-cell activation and inflammatory responses in AD.

The DPP4 gene, which produces a protein identical to the T cell activation antigen CD26 is expressed on the surface of resting and activated T cells, activated B cells, and activated NK cells[192]. It has been implicated in: activation of regulatory T cells, T cell cycle arrest, inhibition synthesis and proliferation of T cells and suppression of inflammatory cytokines by T cells[191]. In this study it was down-regulated in cAD tissues. This is similar to findings in human microarray[178] and flow cytometery studies[190]. Reduced DPP4 expression could lead to less effective regulation of inflammation[192], delayed wound healing[193] and ultimately lesion formation in AD.

PKP2 was slightly up-regulated in lesional and non-lesional skin, and expression correlated with CADESI-03. This was in contrast to the reduced expression reported in a human microarray, although this studied cultured fibroblasts and not whole skin[178], as in this study. Increased PKP2 expression up-regulates beta catenin/T cell factor signalling activity[179], enhancing T-cell survival[180]. Enhanced T cell survival and activity is thought to be a key factor in human and canine AD[31;279].

PPARγ expression was down-regulated in atopic lesional dog skin but not non-lesional skin. Similar findings were seen in a human microarray study of atopic skin[178]. Other members of the PPAR family (PPARβ/δ and PPARα) may represent potential pathogenic candidates as they too have been associated with inflammation and skin barrier function, and have been implicated in the pathogenesis of hAD[209]. PPARγ may be important in the development of chronic inflammation as it is an anti-inflammatory molecule. However, reduced expression of PPARγ is also associated with reduced lamellar body formation and lipid processing, and increased epidermal permeability. The epidermal barrier is of increasing

significance in AD research, particularly following the association of two independent loss of function Filaggrin variants (R501X and 2282del4) with hAD[52]. Filaggrin was considered for investigation but not included in this study due to a lack of reliable primers to quantify its expression in canine skin (data not shown).

SPINK5 was up-regulated in lesional atopic skin compared to healthy skin (fold change 4.68), and its expression correlated with the number of positive IDT reactions. SPINK5, which has been associated with hAD[126;127], regulates hair and skin morphogenesis[170], particularly proteolysis during epithelial formation and keratinocyte differentiation, suggesting altered SPINK5 expression will affect skin barrier function[237]. SPINK5 is also expressed in the thymus, where altered expression has been hypothesised to cause abnormal maturation of T-lymphocytes and Th2 polarisation[280]. SPINK5 protein, in contrast, is down-regulated when associated with atopic symptoms in Netherton Syndrome in humans. This apparent contradiction in expression levels between humans and dogs may reflect differences in pathogenesis, SPINK5 activity in different disease states, and/or poor correlation between RNA and protein expression[281]. It is possible that errors in translation may lead to low protein expression despite enhanced mRNA levels. MicroRNAs, for example, have been observed to bind to the 3' region of mRNA and prevent translation of the message into protein[282], without degrading mRNA which will still be detected by qPCR. It is, therefore possible that SPINK5 mRNA in AD is up-regulated in an attempt to compensate for the low levels of protein but microRNAs may prevent the translation to protein causing a 'feedback loop'. The lower the protein level the higher the mRNA expression level, suggesting that low SPINK5 protein levels lead to skin barrier dysfunction and therefore more allergens can cross the skin barrier eliciting an inflammatory reaction. This is, however, a speculative argument until SPINK5 protein levels in cAD are quantified and functional experiments on microRNAs in SPINK5 can be conducted.

Expression of SGPL1 (Sphingosine-1-phosphate lyase 1) was increased in both lesional and non-lesional skin, as observed in a human qPCR study[171]. Sphingomyelin metabolism enzymes are highly conserved across species[283], and are likely to have similar functions in dogs and humans. SGPL1 is involved in the skin barrier through regulation of sphingosine 1 phosphate and enhanced catalytic cleavage[232], influencing antimicrobial activity as well as keratinocyte proliferation and differentiation[284].

The pro-inflammatory gene CMA1 (Chymase Precursor- Mast cell protease I) is a chymotryptic serine proteinase that is expressed in mast cells and functions as a pro-inflammatory molecule and is associated with hAD[43;135]. It was significantly correlated with the number of IDT reactions but was not significantly up-regulated in lesional and non-lesional skin. SAA-1 (Serum amyloid A-1 protein Precursor), which is an apolipoprotein that recruits immune cells to inflammatory sites and is regulated by interleukin (IL)-1, IL-6 and TNF-α[184;185] and therefore is associated with acute inflammation[185], has shown a significant correlation with both the number of IDT reactions and CADESI-03. However, it was not statistically significantly dysregulated in cAD. Finally, P-selectin which has been shown to prime leukocyte integrin activation during inflammation[270], did not show a statistically significant change in expression, contradicting a canine immunohistochemistry study which reported an increased expression of P-selectin protein[269]. This may be explained by the fact that P-selectin protein is preferentially bound to eosinophils in AD[285] and eosinophils infiltrate the epidermis in AD[22] giving the impression of increased P-selectin protein expression.

Genes correlated with CADESI-03 scores were pro-inflammatory except PKP2, although PKP2 has been shown to have T-cell related functions through implication with the beta catenin/T cell factor signalling activity[179] and T-cell survival[180]. The correlations with

CADESI-03 suggest an association between specific gene activity and disease severity. The strength of the correlations could have been affected by the use of different clinicians in CADESI-03 assessments, although this has been shown to have high inter- and intra-observer reliability[153].

Genes correlated with the IDT results were also mainly inflammatory related, with the exception of SPINK5, although this also has T-cell related activity. The number of positive reactions to IDT was used as a measure of allergen specific IgE because, unlike humans, total IgE levels do not correlate with atopic status or clinical signs, and the specificity and sensitivity of IgE specific serology is variable in dogs[33;276]. In this study the majority of dogs showed positive reactions to the same 5 allergens, these were: *Dermatophagoides farinae*, *D. pteronyssinus*, *Acarus siro*, *Tyrophagus putresceantiae*, *Lepidoglyphus destructor*. These are all house dust mites or storage mites, it has been observed that there is extensive cross-reactivity among house dust and storage mites which may lead to false positive results on an IDT[286].

There was no correlation between CADESI-03 and number of positive reactions on IDT (data not shown; r=0.29 p=0.26), suggesting that IDTs are a measure of individual susceptibility and exposure to allergens but not AD severity. This supports previous evidence that allergic sensitisation is not the only cause of AD; if it were there should be a linear relationship between the number of allergens sensitized too and the severity of disease. The IDT does only assess the reactions to 54 allergens therefore it could be argued that the dog may be allergic to allergens not included on the IDT panel. Even though an animal is sensitized to allergen it does not necessarily mean it comes into contact with those allergens, therefore the symptoms may be mild due to the lack of environmental exposure. Thus, the clinical significance of the associations with the number of positive IDT reactions is unclear

and further investigation of the clinical relevance of IDT reactions to disease severity may be needed.

Human AD studies have shown that some patients with negative skin prick tests have positive specific IgE, and, when given the allergen orally, flare ups were seen at lesional sites[287]. This supports the hypothesis that there is a gut borne and/or dietary component in AD, which is also seen in some canine cases[288]. The atopy patch testing (APT) "may detect relevant sensitisation in the absence of specific IgE"[289]. Comparisons between APT and skin prick test (SPT) demonstrated this; atopic patients showed two patterns of allergic response, one IgE mediated (positive SPT) and one cell mediated (positive APT)[290]. The lack of correlation between CADESI-03 and IDT reactions suggest that non-IgE mediated mechanisms are involved in cAD pathogenesis. Further studies analysing gene expression, CADESI-03, and APT and IDT reactions would help clarify this issue.

A weakness of this study was the inclusion of multiple dog breeds and the variable time scale of the lesions at presentation. This was dictated by the nature of obtaining samples from clinical cases. It is possible, however, that different breeds may have different genetic traits that result in clinical AD, and therefore it would be valuable to study expression and polymorphisms of the genes identified in this study with larger single breed cohorts. It would also be useful to repeat this work on samples taken at defined time points following allergen exposure in experimental canine models of AD. The results may have been influenced by the differing home environments for each dog. The effect of the environment, including diet, is recognised in both canine and human AD[288;291], though not assessed or controlled for in this study (apart from exclusion of food allergy by a 6 week food trial performed as part of the diagnosis).

In conclusion, this study identified 11 genes that were significantly dysregulated in cAD. They are either associated or dysregulated in hAD, or regulate immune reactions and skin barrier function. It is therefore likely that dogs with naturally occurring AD are models for the human condition. Moreover, this gene expression study supports the selection of the identified genes for a large-scale SNP study which will help to elucidate their roles in the pathogenesis of AD.

Chapter 6

SNP Discovery Using Denaturing High-performance Liquid Chromatography and Investigation of a Novel microsatellite in CMA1

6. SNP Discovery Using Denaturing High-performance Liquid Chromatography and Investigation of a Novel Microsatellite in CMA1

6.1 Introduction

Altered expression of candidate genes has been demonstrated in cAD in Chapters 3 and 5, and a number of literature based candidates from Chapter 3 have also been implicated in the pathogenesis of cAD. As this was a candidate gene study with the ultimate aim of identifying disease susceptibility alleles for cAD, the next step was to identify SNPs within the candidate genes for further study. Once identified these SNPs could be included in a large-scale case-control comparison study which would genotype the SNPs and compare frequencies between cases and controls. If a SNP is present in a significantly higher frequency in one group it may be associated with that disease phenotype (see Chapter 1, Figure 1.3).

SNP identification can be done electronically "in silico" by searching databases, such as the Broad Institute or Ensembl, which provide data on SNPs identified by other scientists. However, as yet these canine databases are not well populated and there are a large number SNPs yet to be identified in the canine genome. As such, it was necessary to expand the number of SNPs to investigate in this study by first performing a SNP discovery programme for genes of potential interest in cAD. There are many technologies available for the identification of novel SNPs, the most reliable method is nucleotide sequencing[292]. However, this is an expensive approach and is 'not compatible with cost-effective large-scale population-based genetic screening'[293].

Although DNA sequencing has become more affordable, the cost of re-sequencing for large-scale SNP discovery studies remains high. This means that pre-screening is often used in SNP discovery studies; this allows SNPs to be rapidly and cheaply identified in a large

population in multiple genes. From this, individuals with potentially new SNPs are selected for sequencing to confirm the location and base pair sequence of the SNP. Pre-screening is the most cost-effective and efficient approach to SNP discovery. There are many methods to pre-screen, and a summary is presented in Table 6.1 (adapted from Suh & Vijg (2005)[293]):

Table 6.1 SNP Discovery Methods

Pre-screening Method	Accuracy	Supporting references
Sequencing	High	[294]
SSCP	Low	[295;296]
CFLP	Medium	[297]
CSGE	Low	[298;299]
DHPLC	Medium	[300;301]
DDGE	High	[302;303]
TDGS	High	[304;305]

Table 6.1: Assuming, no human error, sequencing as the gold standard set at 100%, High =100%, Medium =90-100%, Low = 60-90%. SSCP - Single Strand Conformation Polymorphism, CFLP - Cleavage Fragment Length Polymorphisms, CSGE - Conformation-Sensitive Gel Electrophoresis, DHPLC - Denaturing high-performance liquid chromatography, DDGE - Denaturing gradient gel electrophoresis, TDGS - Two dimensional gene scanning

Single Strand Conformation Polymorphism (SSCP) measures electrophoretic mobility and is based on the rationale that electrophoretic mobility is altered if there is a mutation present. This is due to conformational changes in the DNA molecule, caused by the mutation. Not all mutations, however, lead to changes in the secondary structure and although this is a simple method, multiple electrophoretic conditions need to be used, making it time consuming. Moreover, its sensitivity is between 60-95% and this becomes more unreliable with fragments greater than 250bp[306], making screening of a whole gene unrealistic.

Cleavage Fragment Length Polymorphisms (CFLP) use endonuclease enzymes to cleave the DNA molecule; if there are mutations present the pattern and sizes of the resulting fragments (when run on a gel) are different. CFLP is rapid and allows for analysis of large fragments[307]. However, reproducibility can be an issue as this technique is sensitive to the experimental conditions used.

Conformation-Sensitive Gel Electrophoresis (CSGE) detects SNPs by analysing DNA secondary structure; this technique forms hetero and homo duplexes by heat denaturation and reannealing with wild type and mutant DNA[308] (see Figure 6.1 for full explanation of heteroduplex formation). These homoduplexes and heteroduplexes show differences in electrophoretic mobility. However, it is only sensitive enough when combined with other methods such as SSCP[293].

Figure 6.1 Formation of Homoduplexes and Heteroduplexes

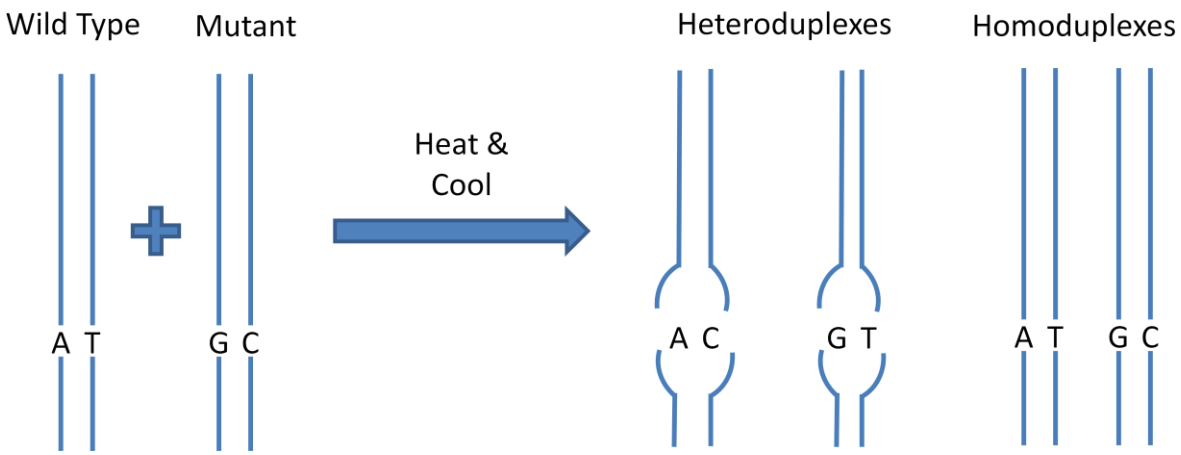


Figure 6.1: The formation of hetero and homoduplexes using wild type and mutant DNA by heating and cooling to reanneal the DNA. Heteroduplexes are formed when the reannealed DNA is mismatched - represented by a 'bump' in the DNA strands in the diagram. Homoduplexes are formed when the DNA reanneals correctly and no mismatch occurs. Thus, the 'bump' or change in the secondary DNA structure allows identification of heteroduplexes and therefore mutations for sequencing.

Denaturing gradient gel electrophoresis (DGGE)[309] and temperature gradient gel electrophoresis (TGGE)[310] are both based on the changes of the secondary structure of DNA fragments under denaturing conditions and the fact that fragments with polymorphisms behave differently to those without, resulting in changes in the electrophoretic mobility. DGGE uses an increased concentration of denaturant and TGGE uses increasing temperature to achieve this. The use of GC rich clamps in DGGE has improved sensitivity to 100%[311].

Two dimensional gene scanning (TDGS) is based on DGGE and allows analysis of an entire gene for all possible sequence variants in one gel under one set of conditions[312;313]. 2D electrophoresis can be combined with multi-colour fluorescent detection[314], fragments can be analysed in parallel and multiplex PCR can be utilised with up to 26 fragments in one reaction, making TDGS a cost-effective alternative to nucleotide sequencing and with an equal accuracy[305].

Denaturing high-performance liquid chromatography (DHPLC)[315;316] uses ion pair reverse phase liquid chromatography to detect heteroduplexes, which are formed as shown in Figure 6.1. DHPLC uses partially denaturing conditions and an acetonitrile gradient to denature the DNA fragments being analysed. If a mutation is present the DNA fragment forms a heteroduplex and is more easily denatured – this means that the retention time in the column is reduced compared to homoduplexes. Therefore mutations can be identified in specific individuals. This approach is advantageous because it is relatively hands free compared to the previously described gel based techniques. It allows fragments between 200 – 700bp to be analysed and is also relatively rapid. However, only one sample can be processed at a time making high throughput processing slow. In Table 6.1, DHPLC is described as medium accuracy; this is because of the difficulty of optimising the temperatures for heteroduplex detection. In DHPLC, multiple runs at different temperatures are required to

ensure that the whole fragment is denatured and all the possible heteroduplexes are identified. If the method includes all the temperatures required for full denaturing of the entire fragment, then the accuracy of DHPLC is 100%. However, if one temperature is missed the accuracy drops to 90%.

TDGS is considered the best technique for studying large fragments and multiple genes. However, in this study the transgenomic WAVE system (which uses DHPLC) was used for the detection of SNPs. This is because only eight genes were analysed, making DHPLC the most cost effective and time efficient approach for this number of genes. The genes (CMA1, SGPL1, DPP4, SPRR1B, S100A8, IL1RL1, MS4A2 and P-selectin) are all candidates for cAD (See Chapter 3) and were selected for SNP discovery because there is little or no SNP information available within these genes in the Broad institute and Ensembl databases. After pre-screening with WAVE, sequencing was performed to characterise the identified SNPs.

It was hoped that by identifying SNPs in these genes they would be included on a large case-control comparison of SNP genotypes which may help define the genetic basis of cAD. In the course of this SNP discovery study a novel microsatellite was identified by subsequent sequencing of mutations indicated by WAVE. This was not included in a large-scale SNP genotyping study and was investigated for an association with cAD separately; the results of this microsatellite association study are presented in this chapter.

Microsatellites or variable number tandem repeats (VNTRs) are polymorphic loci which consist of repeating units of 1-6 base pairs in length. Microsatellites are can be used for identity matching and paternity identification. They have also been implicated in disease pathogenesis through both loss-of-function and gain-of function mechanisms. For example almost 20 unstable repeats (microsatellites) have been associated with different neurological disorders[317]. The most well known of these is Huntington's disease, although there are

several others, including fragile X syndrome (FRAXA)[318], and X-linked spinal and bulbar muscle atrophy (Kennedy's disease) which also have microsatellite associations with disease pathology[319]. Moreover, disease risk can be associated with microsatellites; for example, breast cancer risk in North Indians is associated with microsatellites in the vitamin D receptor[320].

The novel microsatellite was identified in CMA1 which produces the protein mast cell chymase 1. CMA1 is a chymotryptic serine protease which is expressed in mast cells and functions as a pro-inflammatory molecule. It has also been shown to be in an area of linkage with hAD[43;194] and has shown associations with hAD and/or total serum IgE[132-136]. Therefore, it was an excellent candidate for cAD susceptibility and the discovery of a novel microsatellite offered an opportunity to investigate a potential association with different kind of mutation (other than SNPs) in a large number of cases and controls.

6.2 Methods

WAVE analysis

Sample selection

Blood samples from 11 Boxers and 11 West Highland White Terriers (WHWTs), excess to diagnostic requirements and supplied by the Veterinary Laboratory agency (VLA) were selected for this experiment. The samples represented a random population of dogs with no disease phenotype information. Because this was a SNP discovery exercise, the use of a diverse random population was recommended to obtain rare SNPs. The blood was collected, extracted and processed as set out in Chapter 2. The DNA was normalized to 2ng/µl with TE for this experiment.

Primer design

Primers were designed using the genomic sequences of candidate genes; the focus was on promoter regions and exons as these are coding regions and are more likely to contain SNPs that can affect disease pathology.

Primers were designed using primer3 plus software[321]; highly repetitive areas were avoided and the primers were designed around areas between 200 – 700bp. To reduce self complementary and primer dimer reactions, the maximum self complementary was reduced to 4, the maximum self 3 primed was set to 2, and the maximum poly x was set at 4. Temperatures for primer design were set to a minimum 57°C, optimum of 60°C and maximum of 63°C. A GC clamp was added to the primer design parameters to increase stability of binding.

The primers were ordered from Metabion (Germany), 0.02mmol scale and standard purification (see Table 6.2 for the primer sequences).

Table 6.2 Primers for WAVE

Gene	Location	Forward	Reverse	Product size
CMA1	Promoter	cagattcccccagtcacaag	taacaggcagagtgtcgtttc	375
	Exon 5	ttgaaggagaaggccaacc	tggtcaaaagccatgtagtgtc	203
	Exon 6	ctcttctgtgtgctggggtag	gggtctgagggatgtagtgg	250
SGPL1	Exon 2	cattgaagaagtgagtgaagaacag	cccattaaactcgcctgaag	206
	Exon 5	gtgctctcacctgtgtgcttc	gccgaaccagtcctccac	294
	Exon 14	accgccagtaacgacctc	tcagggtcacagagtatccag	197
DPP4	promoter	acaaacaaacagcaacccaag	ttccacaccccacactcac	493
	Promoter	gtgtggaaggctgaggtaatg	cttgaacttggagggtggtg	278
	Exon 7	caagcactcctcatttgatgc	tgaccttcggattcatctcc	287
	Exon 21	caagcactcctcatttgatgc	tgaccttcggattcatctcc	287
SPRR1B	Promoter	cacttctcccccaacacaag	cctcaccagccctcttacac	414
	Promoter	ccctttcagcaacaccactc	gaccatgtggcatacctgtg	434
	Exon 1	agaagcaaccctgcatcc	gtcactgttgagggacatgg	214
	Exon 3	caacaggtgaaacagacttgc	aggtcatggctaaggaccac	300
S100A8	Exon 1 - 2	cccatgtgtttcgtgttgc	tattattcggcagggacagc	500
	Promoter	gggattagcagcagcaagag	tccagacctcccttagctttc	480
IL1RL1	Exon 10	cacagcaacaagaaatctgacc	cggaaatctggtgcattagc	461
	Promoter	tgccaaatgaagtgttgagg	ccaaactgaaaggcagaagg	459
	Promoter	ttggtgagtcttggggacac	aggttgatggtgcctcctg	573
MS4A2	Promoter	aagatggcacgggtttaatg	tgattttgacgatggcagag	419
	Exon 5-6	acacacagggctattgaacg	gcttcagcttgtcattcagg	700
P-SELECTIN	Exon 5	aagggctcatgctattgctc	ggccactgtcatctgtactcac	296
	Exon 11	gactctgttatgctccgtttattg	gcctgttcctttcccattac	289

Touchdown PCR, WAVE PCR & WAVE assay design

After Touchdown PCR (see Chapter 2 section 2.4; reagent volumes used per well (µl): Taq 0.25, dNTP 0.5, Primer F 0.25, Primer R 0.25, X10 2.5, dH20 16.25) the product was run on a DNA engine Dyad™ peltier thermal cycler or a DNA engine Tetrad™ PTC-225 peltier thermal cycler for a short denaturing program to form heteroduplexes for analysis (see Figure 6.1). The denaturing program was as follows: 95ºC 5 minutes, then reduce temp by 1.5 ºC per minute until 25 ºC was reached.

The melting temperatures and percentage of buffers were selected by the WAVEMAKER V4.1.44 (Transgenomic, Huston, Texas, USA) software by uploading sequences and generating methods. Occasionally more than one temperature was required to analyse the full sequence (see Appendix 3 for full details on temperatures and fragments uploaded). The temperatures and percentage buffers could be altered manually to optimize the resolution of the chromatographs.

The WAVE was equilibrated before use with 50% buffer A (Transgenomic Bioconsumables, Huston, Texas, USA) and 50% buffer B (Transgenomic Bioconsumables, Huston, Texas, USA) for 15 minutes at a flow rate of 1.5/min to‘clean’ the column before use. Low range mutation standard (Transgenomic Bioconsumables, Huston, Texas, USA) was used to check the efficacy of the column before, during and after the run. 5µl of sample was run through the WAVE at the predefined temperature generating a chromatograph. Before each temperature/method change, blanks were run to ensure the column was at the correct temperature and equilibrated therefore reducing false positive results. Between plates a 75% acetonitrile (Transgenomic Bioconsumables, Huston, Texas, USA) wash was performed to ensure no product was left on the column before starting the next plate.

Chromatograph analysis

The chromatographs from the WAVE were analysed for double or unusual peaks and then chosen for sequencing.

Sequencing

The samples were prepared for sequencing by taking 2µl of PCR product and adding 7µl of distilled water and 1µl Exo-SAPIT® (USB, USA). A SAP purification program was run on the Dyad™ peltier thermal cycler or a DNA engine Tetrad™ PTC-225 peltier thermal cycler (MJ research Inc., Waltham MA, USA). The conditions were as follows: 37ºC for 1 hour followed by 85ºC for 15 minutes.

Following this 2µl of cleaned PCR product were aliquotted into a 96 well plate, to this 8µl of forward OR reverse primers at 0.4mM were added to each well (the primers used were not specific sequencing primers, the original PCR primers were re-used).

Sequencing was performed by the DNA sequencing facility in the Faculty of Life Sciences at the University of Manchester, using the Applied Biosystems 3730 DNA Analyzer (Applied biosystems, Foster City, CA, USA).

Microsatellite analysis

Sample selection

Disease diagnosis, sample collection and processing of DNA samples from dogs with and without cAD were performed detailed in Chapter 2. In total, 659 DNA samples were included on the microsatellite array. Eight breeds were included in this study; Table 6.3 shows the numbers of each breed and disease status.

Table 6.3 Samples included in Microsatellite genotyping

Breed	Cases	Controls	Total
Boxer	11	24	35
GSD	19	38	57
Labrador	64	129	193
G. Retriever	40	66	106
Shiba Inu	23	33	56
Shih Tzu	47	62	109
Pit Bull	20	17	37
WHWT	18	48	66
Totals	**242**	**417**	**659**

Primer design

The microsatellite was discovered in the promoter region of CMA1 (see WAVE results section, Figure 6.3 for details). The primers were designed as described in the WAVE analysis primer design section above. The primers were supplied by Metabion (Germany), at 0.02mmol scale with standard purification. The forward primers 5' end was labelled with fluorescent probe FAM for identification of the microsatellite size:

Forward: CTGGGAGAAAGCTCCATGTC

Reverse: GGCAGAGTGTCGTTTCTCTTG

Assay design

The initial PCR reaction to amplify the microsatellite was performed as set out in Chapter 2. After the initial PCR, 1µl of PCR product was aliquotted into a 384 well plate and a mix

(4.8µl/0.2µl) of de-ionised Formamide (Sigma-Aldrich Inc.) and ABI prism® Rox400 (Applied Biosystems, Foster City, CA, USA) was added into each well giving a reaction volume of 6µl. The ABI prism® 3100 genetic analyzer data collection software V1.1 (Applied biosystems, Foster City, CA, USA) was used to set up the run; the following variables were selected - dye set D, default POP4 run module which uses ABI prism® 3100 POP4™ polymer (Applied Biosystems, Foster City, CA, USA) and GS400HD analysis module as per the manufacturer's instructions. The ABI prism® Genescan software V3.7 (Applied biosystems, Foster City, CA, USA) was then used to analyse the data according to the ROX size standard used, the files were saved from this and used in the ABI prism® Genotyper software V3.7 (Applied biosystems, Foster City, CA, USA). Genotyper allows accurate sizing between plates and wells, using macros to group the results into genotypes/microsatellites which are dependent on the length of the microsatellite.

Statistical analysis

Gene frequencies and phenotype frequencies were estimated by direct counting. The percentage distribution of the genotype in cases and controls was calculated in order to identify any differences. The significance of the distribution of genotype between cases and controls was tested using 2x2 contigency tables; the chi –squared ($\chi 2$) test with Yates' correction was used to produce p values. The statistical tests were only performed on genotypes which were present in more than 5 dogs to ensure statistical robustness. These analyses were performed on the whole data set (all breeds as one panel) and within individual breeds to test the association of microsatellite genotype with cAD. Further to this analysis, the association between breed and microsatellite genotype was also tested using $\chi 2$ analysis.

6.3 Results

WAVE Results

To analyse the results from WAVE, chromatographs were generated to display the retention time on the WAVE column. As described in the introduction, a change in retention time suggests a SNP is present. Heteroduplexes (mutations) come off the column first as they are more easily denatured; this means they have a lower retention time than homoduplexes. Therefore, when looking at WAVE chromatographs it is possible to either get a single peak after the size standard (Figure 6.2), which means there are no SNPs, only homoduplexes, or a double peak after the size standard indicating a heteroduplex, suggesting a SNP is present (Figure 6.2). Figure 6.2 demonstrates the differences in chromatographs when homo and heteroduplexes are present.

Figure 6.2 Retention Time Graphs (Chromatographs) from WAVE

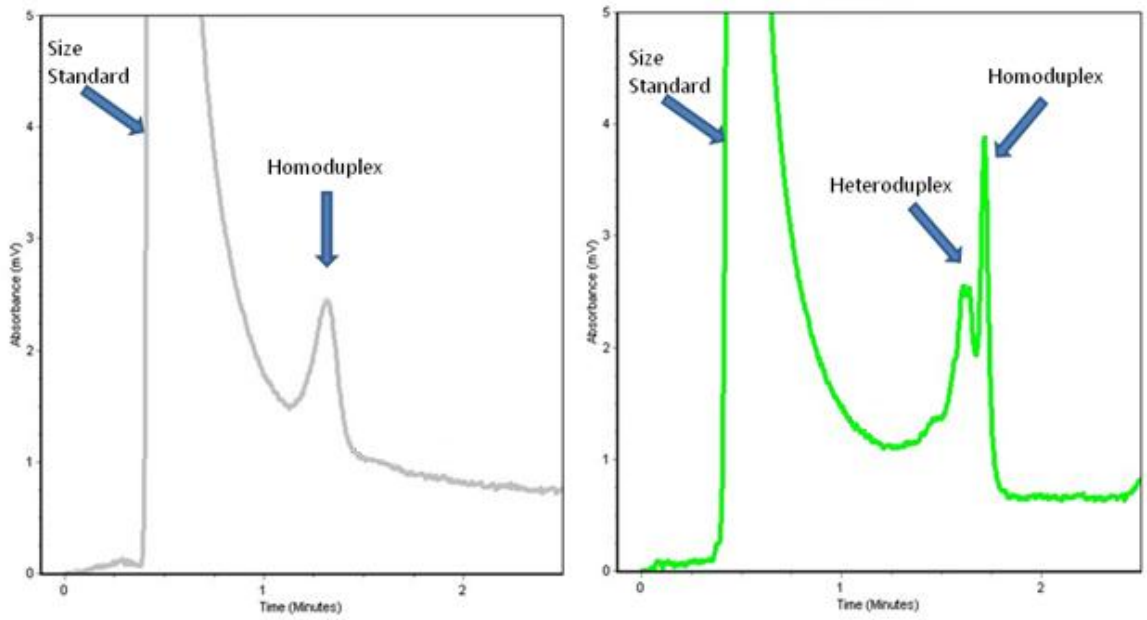


Figure 6.2: Two examples of chromatographs generated by the WAVE. The first chromatograph displays a negative SNP result showing a single peak indicating that only homoduplexes are present. The second graph shows two peaks indicating that both homo and heteroduplexes are present indicating a SNP may be present in this individual.

Figure 6.3 shows the chromatograph results for this study, but only when a SNP was identified. Potential SNPs were identified in MS4A2 exons 5 and 6 (Figure 6.3a), S100A8 promoter regions (Figure 6.3b & c), SPRR1B promoter region (Figure 6.3d), CMA1 promoter region (Figure 6.3e), SPRR1B exon 3(Figure 6.3f &g), and SGPL1 exon 2 (Figure 6.3h).

Figure 6.3 Chromatograph examples of individual dogs with SNPs

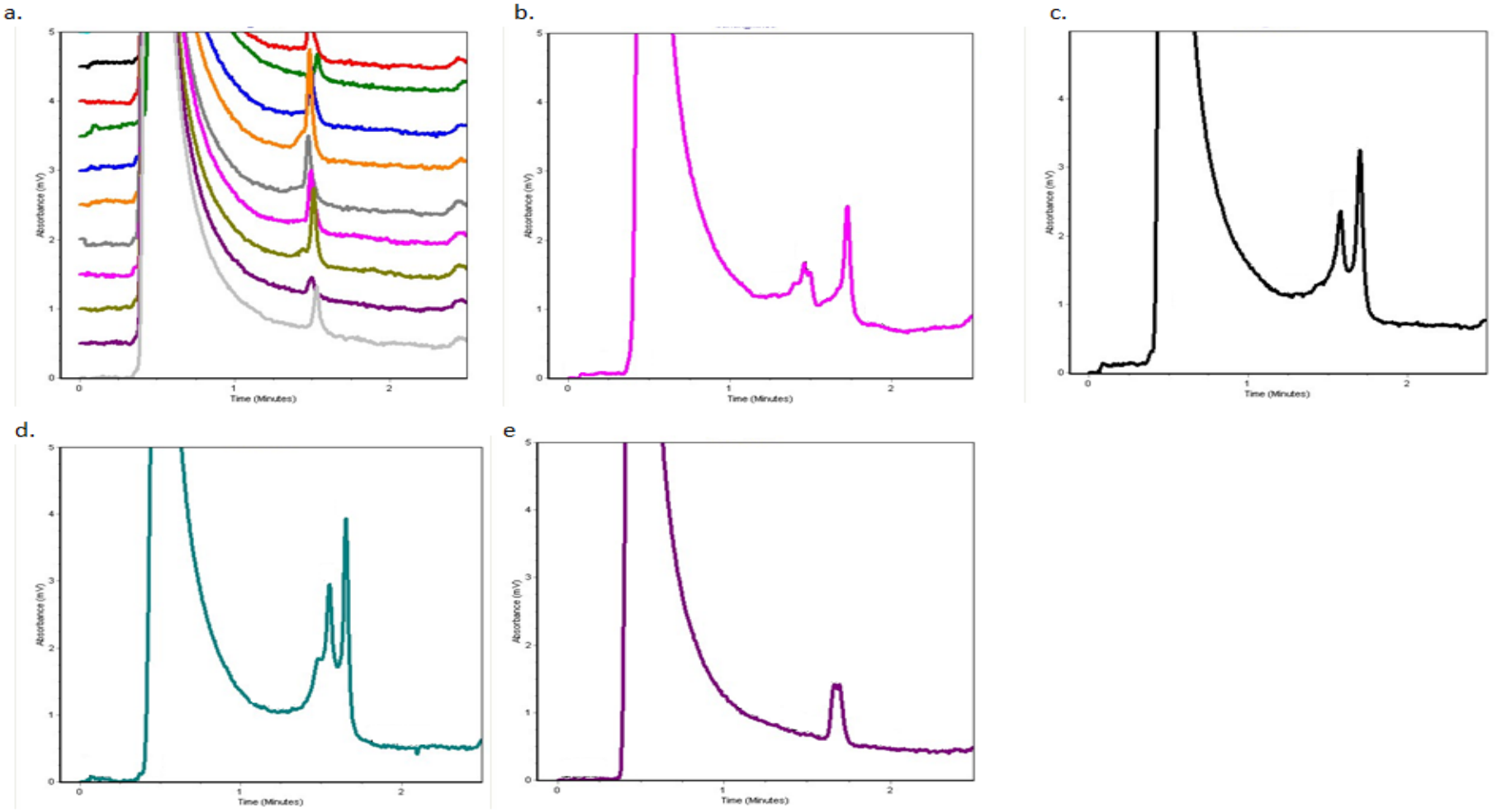


Figure 6.3 cont.

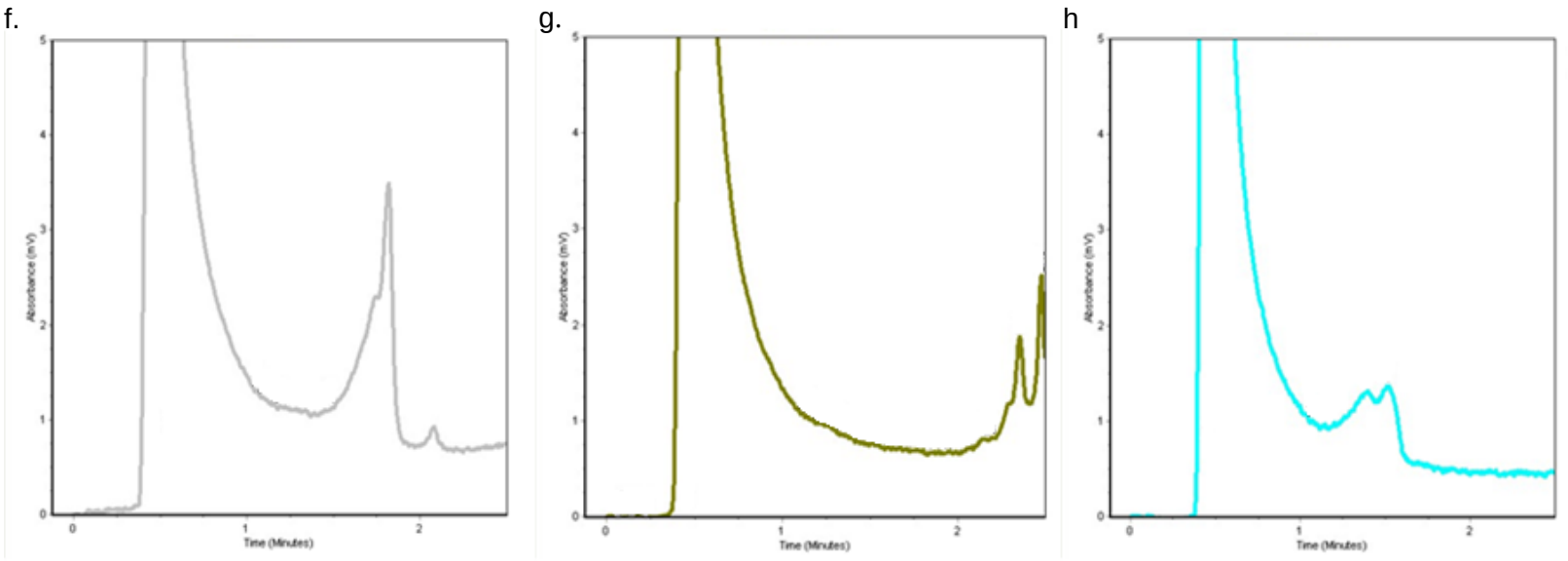


Figure 6.3: (a) MS4A2 exon 5 and 6 shows a shift in the position of the peaks (change in retention time); this suggests that there is a SNP present. (b) S100A8 promoter region showed three distinct peaks, suggesting multiple SNPs were present. (c) S100A8 promoter region shows two peaks suggesting that different SNPs are present in this individual compared to b (d) SPRR1B promoter region (e) CMA1 promoter region has a flattening of the top of the peak suggesting the presence of two peaks and therefore a SNP (f) SPRR1B exon 3 shows a large peak with a shoulder which could be a distinct peak or a PCR artifact; there is also a small peak to the right which could indicate the presence of a SNP. (g) SPRR1B exon 3 shows two peaks in a different position to (f), suggesting that there are different SNPs present in this individual. (h) SGPL1 exon 2 shows two small peaks suggesting a SNP is present.

Sequencing

The individuals identified by WAVE as having SNPs present were selected for sequencing to characterize the SNPs. For all sequences from each individual see Appendix 4. A summary of the SNPs found in this study is shown in Table 6.4. For the position of each SNP in the gene see Appendix 5.

In all, 11 novel SNPs and 1 novel microsatellite were found by WAVE in the 8 genes under analysis. Three existing database (Ensembl) SNPs were confirmed by WAVE.

Table 6.4 Numbers of SNPs identified by WAVE

Gene	Region	Novel SNPs	Ensembl SNPs
CMA1	Promoter	1*	
MS4A2	Exon 5 & 6	3	2
S100A8	Promoter	3	1
SGPL1	Exon 2	1	
SPRR1B	Exon 3	2	
SPRR1B	Promoter	2	

* Denotes a microsatellite instead of a SNP

Microsatellite results

The microsatellite identified by WAVE and sequencing was genotyped in 659 dogs (64 excluded from the analysis due to low genotyping success) across 8 breeds. The microsatellite analysis identified 28 different genotypes, 26 of which were found in the atopic dogs and 24 in the healthy controls. For the most part the numbers of individuals with each of these 28 genotypes was low. The exception to this was A06/A06, which was by far the most common microsatellite genotype with over half the cases and controls typed having this

genotype. For a full list of allele frequencies see Appendix 6. This also shows the relative sizes of each allele. The results below are divided into case-control comparison across all 8 breeds and an individual case-control comparison in each breed.

Case-Control Comparison across All Breeds

Microsatellite genotypes which were found in less than 5 dogs (in either cases or controls) were excluded from $\chi 2$ analysis; this is because there was not enough data to produce an accurate analysis. On this basis, 20 genotypes were excluded from analysis, with the remaining 8 genotypes displayed in Table 6.5 (this shows the p value results from the $\chi 2$ test and the percentage distribution of alleles in the case-control comparison across all 8 breeds). None of the tested alleles shows a significant difference in prevalence between cases and controls.

Table 6.5 Case-Control Analysis in All Breeds: Chi Squared Test Results for Microsatellite Genotypes found in CMA1

Allele 1	Allele 2	Number of Cases	% Cases	Number of Controls	% Controls	p value
A10	A10	8	3.6	25	6.7	0.15
A06	A07	6	2.7	19	5.1	0.22
A04	A06	10	4.5	9	2.4	0.26
A06	A13	9	4.0	20	5.4	0.58
A01	A06	5	2.2	6	1.6	0.82
A10	A13	5	2.2	6	1.6	0.82
A06	A10	26	11.6	41	11.1	0.94
A06	A06	125	55.8	209	56.3	0.97

Case-Control Comparison: Individual Breeds

Each of the eight breeds under study were analysed separately for any significant association between the microsatellite genotype and cAD. As in the previous analysis, microsatellite genotypes which were found in less than 5 dogs (in either cases or controls)

were excluded from $\chi 2$ analysis, meaning that in most cases only the A06/A06 genotype could be analysed. Only one breed, Shiba Inu, showed a significant association between the microsatellite genotype A06/A06 and cAD.

Table 6.6 shows the p value results from the $\chi 2$ test and the percentage distribution for each genotype in the case-control comparison in each individual breed.

Table 6.6 Case-Control Analysis in Individual Breeds: Chi Squared Test Results for Microsatellite Genotypes found in CMA1

Breed	Allele 1	Allele 2	Number of Cases	% cases	Number of Controls	% controls	p value
Shiba Inu	A06	A06	12	54.5	5	15.6	0.006*
Shih Tzu	A06	A06	7	15.2	19	31.1	0.09
WHWT	A06	A06	14	93.3	36	97.3	0.50
Boxer	A06	A06	7	63.6	6	42.9	0.53
German shepherd dog	A06	A06	9	52.9	16	42.1	0.65
Pit Bull terrier	A06	A06	13	68.4	9	56.3	0.70
golden retriever	A06	A06	21	58.3	34	63.0	0.83
Shih Tzu	A06	A10	12	26.1	14	23.0	0.88
Shih Tzu	A10	A10	5	10.9	7	11.5	0.92
Labrador	A06	A06	42	72.4	84	70.6	0.94
Shih Tzu	A06	A13	7	15.2	9	14.8	0.95

* Denotes statistically significant

Further to the case-control analysis, a comparison across breeds was conducted to see whether the variation in microsatellite genotype was attributable to breed. Again, microsatellite genotypes which were found in less than 5 dogs (in any breed) were excluded from $\chi 2$ analysis, leaving only 2 microsatellite genotypes that could be tested. Table 6.7 shows the percentage distribution for each genotype in the breed comparison and the p values from the $\chi 2$ test. It can be clearly seen that in the two microsatellite genotypes tested there was a significant difference between the breeds, suggesting that the microsatellite is associated with breed rather than disease status.

Table 6.7 Breed Comparison of Microsatellite Genotype: Chi Squared Results

Allele 1	Allele 2	%Lab	%Gret	%GSD	%Box	%WHWT	%Pit	%Shi	%Tzu	χ^2 *P* value
A06	A06	71.2	61.1	45.5	52.0	96.2	62.9	31.5	24.3	2.09×10^{-8}
A06	A10	4.5	5.6	18.2	32.0	0.0	11.4	11.1	24.3	6.32×10^{-8}

Lab = Labrador, Gret=Golden retriever, GSD= German shepherd dog, Box= Boxer, WHWT= West Highland white terrier, Pit= Pit Bull terrier, Shi = Shiba Inu, Tzu = Shih Tzu

6.4 Discussion

Eleven novel SNPs and 1 novel microsatellite have been identified in 5 genes by WAVE analysis. These SNPs, however, require further confirmation by screening in a large population to allow the minor allele frequency (MAF) and Hardy-Weinberg equilibrium (HWE) to be estimated. Until this is done, the frequency of these SNPs in the population is unknown. They can be still used in a case-control study, but if one or more are found to be significant further characterization will be necessary to submit the SNP to dbSNP databases.

The SNP discovery experiment could have been improved by including more dogs of different breeds to identify more SNPs across a larger population. It may have been beneficial to use a technique such as TDGS which only requires one experimental condition to identify all SNPs, whereas WAVE requires multiple temperatures and buffer concentrations. This suggests that it is possible that some SNPs were missed.

WAVE analysis can be of use in SNP identification, as it can screen a given number of dogs/breeds in the population to identify novel SNPs in these breeds. WAVE, however, cannot identify whether these SNPs are causative of the disease. These SNPs will therefore

be included on a large-scale case/control comparison of SNP genotypes to investigate potential associations with cAD.

The microsatellite analysis showed that when all the breeds were analysed together, cAD had no significant association with the microsatellite discovered by WAVE. However, in the individual breed analysis the Shiba Inu did show a significant association between the A06/A06 genotype and cAD (odds ratio = 6.48, data not shown). Breed association analysis, nevertheless, appeared to indicate that it was more strongly associated with breed than disease status. There were a number of rare microsatellite alleles that had to be excluded from the analysis as they were found in less than 5 dogs. It may be interesting to follow up disease and breed association in more dogs.

The apparent association of the microsatellite in the Shiba Inu is interesting and is worth further investigation. Only 17 Shiba Inu's were A06/A06, and to validate the potential association with Shiba Inu microsatellite genotype and cAD more dogs of this breed would have to be typed. However, it is likely that the association seen is due to the low numbers of Shiba Inu and the microsatellite genotype is associated with breed rather than disease status.

In conclusion, the novel microsatellite identified in CMA1 has shown a potential association with cAD in the Shiba Inu and a strong association with breed. SNPs were identified by WAVE in five candidate genes and will be included on a large-scale case-control comparison to test for an association with cAD.

Chapter 7

Genome Wide Association Study - Identifying Canine Atopic Dermatitis Susceptibility Genes

7. Genome Wide Association Study - Identifying Canine Atopic Dermatitis Susceptibility Genes

This work is presented in: **Wood, S.H.**, Ke, X., Nuttall, T., McEwan, N., Ollier, W.E., Carter, S.D., (2009) Genome Wide Association Analysis of Canine Atopic Dermatitis and Identification of Disease Related SNPs, *Immunogenetics,* 61(11-12):765-72

7.1 Introduction

In Chapters 3 to 6, candidate genes and SNPs for further investigation of the genetic basis of cAD were identified. For the most part, however, the genes were selected because of literature based evidence of their potential involvement in AD. This chapter introduces a new and unbiased method for the identification of novel SNPs associated with a disease phenotype. This method is referred to as genome wide association study (GWAS) and enables the simultaneous study of thousands of SNPs in individual dogs.

In humans, GWAS has been shown to be an effective approach in identifying gene polymorphisms associated with disease phenotypes. GWAS approaches use simultaneous high density SNP genotyping with the aim of providing coverage across the complete genome. This hypothesis-free approach allows many novel disease susceptibility loci to be identified in a single experiment. Given the chance of false positive discovery through multiple testing, findings should be replicated by other molecular approaches.

The Welcome Trust Case-Control Consortium (WTCCC) is a good example of the usefulness of GWAS. This has recently met with success in identifying novel SNPs in bipolar disorder, coronary heart disease, Rheumatoid arthritis, and type 1 and 2 diabetes in humans[322]. Moreover, GWAS was recently used to investigate hAD in approximately 1000 cases and 1000 controls, and an association was found with an intergenic SNP on

chromosome 11. Validation by another technique in approximately 3000 cases and 4000 controls was required to confirm the association[323]. This demonstrates the need for validation on large numbers of cases and controls (usually mulitple thousand) to achieve sufficient statistical power to detect small genetic risk factors (≤1.5) in human studies. In the recent hAD study, the chance of identifying a risk factor of 1.5 was 84%. This dropped to 12% when identifying a risk factor of 1.2[323]. The most recent SNP arrays used in human studies are of greater density (up to $1.6x10^6$ SNPs), which is needed to detect association effects, as LD across the genome is weak in humans (10-100 kb[324;325]).

Domestic dog breeds are of recent origin and represent highly inbred genetically isolated populations. Consequently, LD extends over long distances (0.8 -5 Mb[107;326]). This is advantageous in GWAS as fewer genetic markers and smaller sample sizes are required to find disease associations in dogs[107;108]. For example, to achieve complete coverage of the dog genome, only 5,000 to 30,000 SNP markers are required[324], compared to an estimated 200,000 to 500,000 SNP markers in humans[108], making GWAS in dogs both cheaper and easier to conduct[107;108].

Previous genome wide genetic studies in canines have mainly been done using microsatellites and have mapped mutations causing lens luxation[327], putative Quantitative Trait Loci (QTLs) in osteoarthritis[328], dilated cardiomyopathy (DCM)[329] and cone-rod dystrophy[330]. One study used a panel of 1500 SNPs in a family of Boxers to identify the loci responsible for white spotting[331], although not truly a genome wide study as the coverage was not complete enough. The only true genome wide SNP based study in dogs used an Illumina custom made array (26,625 SNPs) to identify two monogenetic traits (white spotting[332] and hair ridges in ridgeback dogs[333]), but used only 10 cases and 10 controls[324].

High density SNP arrays for GWAS in dogs are now available and are beginning to be used to analyse both single gene and complex canine diseases. As a tool for genetic studies, these systems represent a huge leap forward in our ability to understand complex diseases with a genetic component. Until their development, it had been necessary to select individual candidate genes for analysis and to sequence the potential SNPs. As each SNP array contains thousands of SNPs, the chances of discovery are much greater than previously possible.

This is the first study to investigate the genetics of cAD using GWAS. The study used the Illumina canine SNP20 chip with 22,362 SNPs assembled from the CanFam2.0 genome release from the Broad Institute based on Boxer and poodle sequences. In the production of this array highly polymorphic loci were chosen from a diverse population of dog breeds. This means that there is likely to be segregation amongst breeds, making the assay suitable for within-breed association studies, and therefore a single breed, the golden retriever, was used.

An issue when analysing GWAS data is that allele differences between cases and controls may be unrelated to the disease phenotype, resulting in false positive associations. Inappropriate statistical correction and genotyping errors can lead to false positive associations. These are taken into account relatively easily by calculating the minor allele frequency (MAF) and Hardy-Weinberg equilibrium (HWE), and applying cut-offs to the data. The Benjamini and Hochberg False Discovery Rate[151] correction is also used to limit false positives arising from inappropriate statistical correction.

The most important factor in false positive discovery is the presence of unrecognised populations structure[334] within the samples used for study. Population structure in this type of study is defined as the presence of a systematic difference in allele frequencies between subpopulations within a study population. The stratification is caused by non-random mating between groups, which in most cases is due to physical separation (e.g., populations of

African and European descent), leading to genetic drift of allele frequencies in each group. Population structure can be a problem for case-control based association studies as associations found may be due to the underlying structure of the population and not a disease associated locus. In addition, the real disease causing locus may not be found if less prevalent in the population of cases chosen for study. There are a number of ways to control for the effects of population structure which compensate for any population bias. The two most common methods are genomic control correction, which is a non-parametric method for controlling the inflation of test statistics[335], and structured association methods[336], which use genetic information to estimate and control for population structure. A structured association method, Identity-By-State (IBS) clustering, was used in this study. Following the IBS clustering, a meta-analysis was performed taking into account the populations identified by IBS and allowing all data to be compared without population structure effects.

This is the first canine study to employ these human-based population structure adjustment approaches, and therefore the "unadjusted" results are also presented for comparison (See Figure 7.1 for the methodology).

Until now the genetics and immunopathogenesis of cAD has been largely limited to cytokine profiling and gene expression studies[31;274;337]. It was hoped that the results of this study would help identify novel loci in cAD, which could be validated at a later stage in large-scale case-control comparison studies. These data may identify possible novel pathogenic pathways and potential routes for therapeutic intervention. This form of analysis should narrow down the search for cAD susceptibility genes.

7.2 Methods

Sample selection for whole genome scan

Forty eight golden retrievers were selected for GWAS. 25 dogs with AD (14 female, 11 male, age range: 2 to 13 years,) and 23 controls (9 female, 14 male, age range: 1-14 years) were included. Samples were collected by the WALTHAM® Centre for Pet Nutrition from Japanese and US veterinary clinics. The diagnoses of cAD and control selection were performed as stated in Chapter 2.1. Blood samples were taken as described in Chapter 2.2 and processed according to Chapter 2.3. The DNA was normalised to 50ng/µl.

GWAS Array

The SNP array analysis of the 48 golden retriever samples was performed using the Illumina canine SNP20 array (San Diego, California), by the Barts and the London Genome Centre according to the manufacturer's instructions using Illumina's software to call the data.

GWAS data were analysed using the freely available whole genome association analysis toolset; PLINK[338]. Only SNPs which conformed to Hardy-Weinberg expectations, exhibited 90% genotyping success and had a minor allele frequency ≥0.01 were included in subsequent analysis. Two hundred and twenty-nine SNPs were excluded on this basis.

Figure 7.1 shows the methodology for this study and explains the separate analysis of "population structure adjusted" and "unadjusted" results.

Figure 7.1 Study Methodology and Analysis

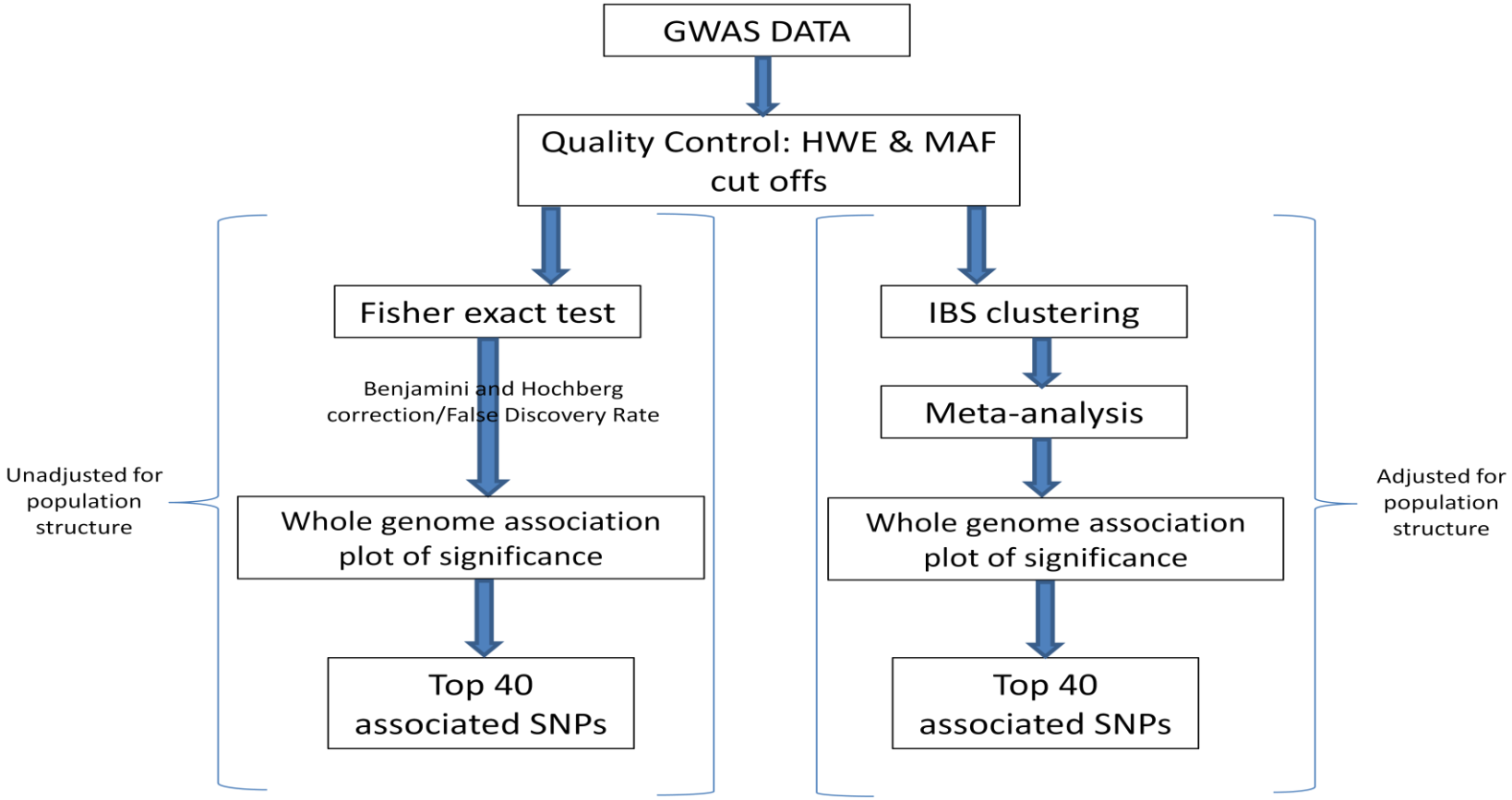


Figure 7.1: The GWAS data was first subject to quality control procedures described in the main text. Hardy-Weinberg Equilibrium (HWE) and minor allele frequency (MAF) cut offs were applied to the data. Two analyses were performed: 1. Fisher exact test and Benjimini and Hochberg correction ("unadjusted" for population structure), and, 2. IBS clustering and meta-analysis ("Adjusted" for population structure). Both analyses generated a whole genome association plot of significance and the top 40 associated SNPs from each analysis are presented.

As shown in Figure. 7.1, comparisions between cases and controls without accounting for population structure, using the Fisher exact test were performed and the Benjamini and Hochberg correction/False Discovery Rate was applied[151]. Significance values from this were used to generate a whole genome association plot of significance (also known as a Manhattan plot) using the statistical package R[339] (the R script was written by Matthew Settles of Washington State of University). This displays the results for each chromosome as a log of the p value. Further to this a Q-Q plot was generated to show the difference between observed and expected results.

To account for the effects of population structure the GWAS results were re-analysed. Population structure was defined using IBS clustering methods on the GWAS data, allowing differences in genotype across the array to be calculated and individuals grouped according to these differences (see Figure.7.1). Further to this Multi-Dimensional Scaling (MDS) was used to visually represent the population clusters. Using the clusters defined by IBS, a meta-analysis (Cochran-Mantel-Haenszel test for 2x2xK stratified tables) was used to compare all results in the cases and controls, independent of population structure. A whole genome association plot of significance was generated for these population structure adjusted results. Using a HWE cut off of 0.05 and a 95% confidence interval the top 40 SNPs from both analyses were identified.

Haplotype analysis

PLINK was used to perform haplotype analysis on the GWAS data. Haplotype association tests were performed using a sliding window specification. Originally a sliding window of 3 SNPs was specified, but after this analysis the sliding window was dropped to 2 SNPs. These data were subject to the same quality control procedures specified above. The p values from the haplotype analysis were compared to those of the single SNP associations. If

the p value is lower in the haplotype analysis compared to the single SNP analysis it indicates that the haplotype is significant. Using Ensembl[159;160] chromosome view the haplotype positions were represented visually.

7.3 Results

Figure 7.2 shows the "unadjusted" results from the Fisher exact test as a whole genome association plot of significance for individual SNPs. This shows the log of the p value, and results above the cut off of 3 are considered relevant and significant. The 40 most significant SNPs from the Fisher exact test are listed in Table 7.1 (all of which are above the cut off of 3 on the genome association plot).

Figure 7.3 shows the Q-Q plot, which displays the difference between the observed and expected results. The large difference between the results suggests that there are wide differences in population structure. The genomic control inflation factor quantifies this difference and can be used to adjust the "population structure unadjusted" p values, to attempt to account for population structure. However, it is considered that IBS clustering and MDS plots do this with greater accuracy (see Figure 7.4).

Figure 7.2 Population Structure Unadjusted: Whole genome association plot of significance for Canine Atopic Dermatitis

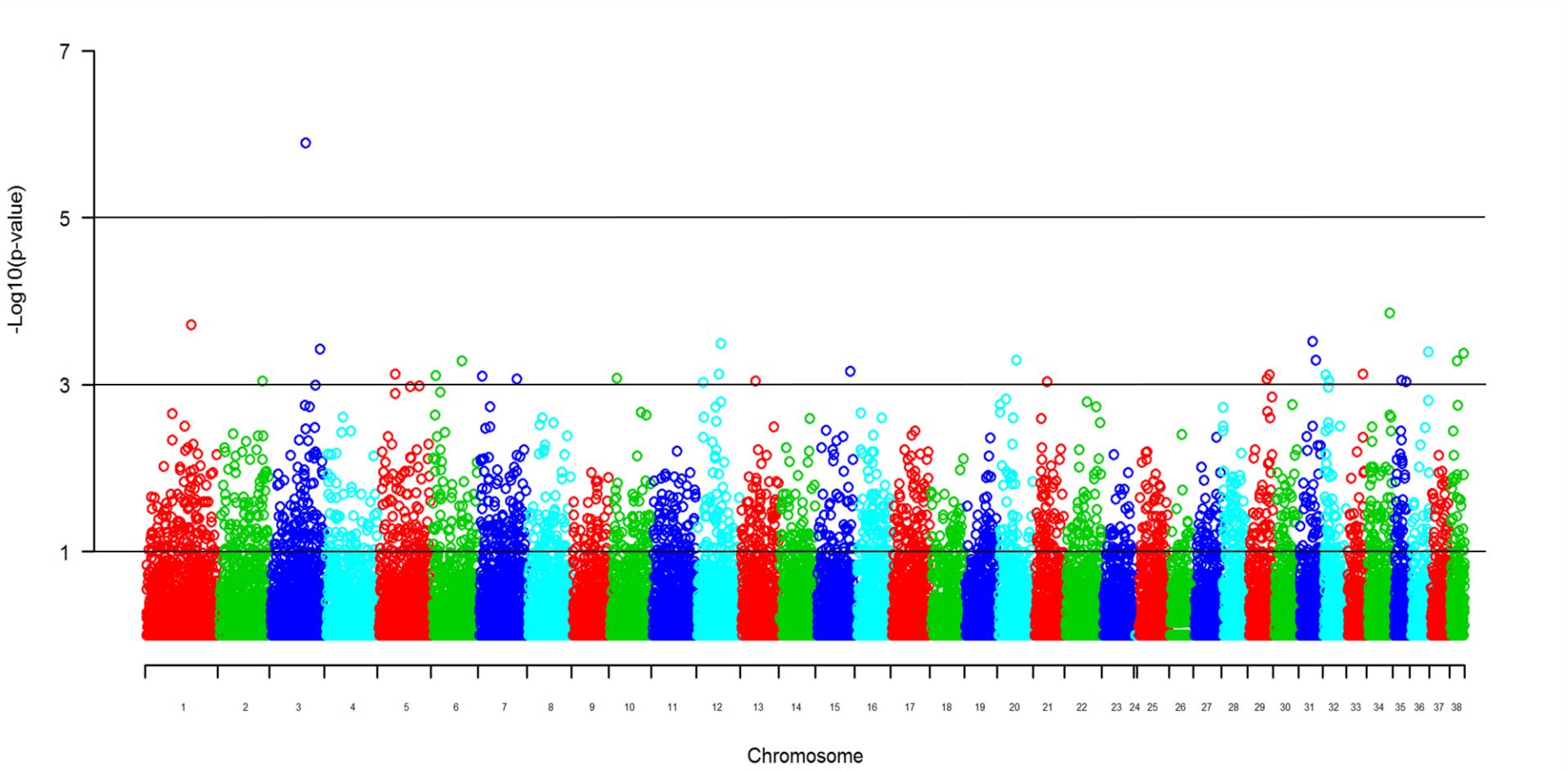


Figure 7.2: A plot showing significance of association of all SNPs in the GWAS using Fisher's exact test ("unadjusted" for population structure) with atopic and healthy control phenotypes. SNPs are plotted on the x-axis according to their position on each chromosome against association with atopic dermatitis on the y-axis (shown as -log10 P value).

Figure 7.3 Q-Q plot of the observed and expected results from the GWAS

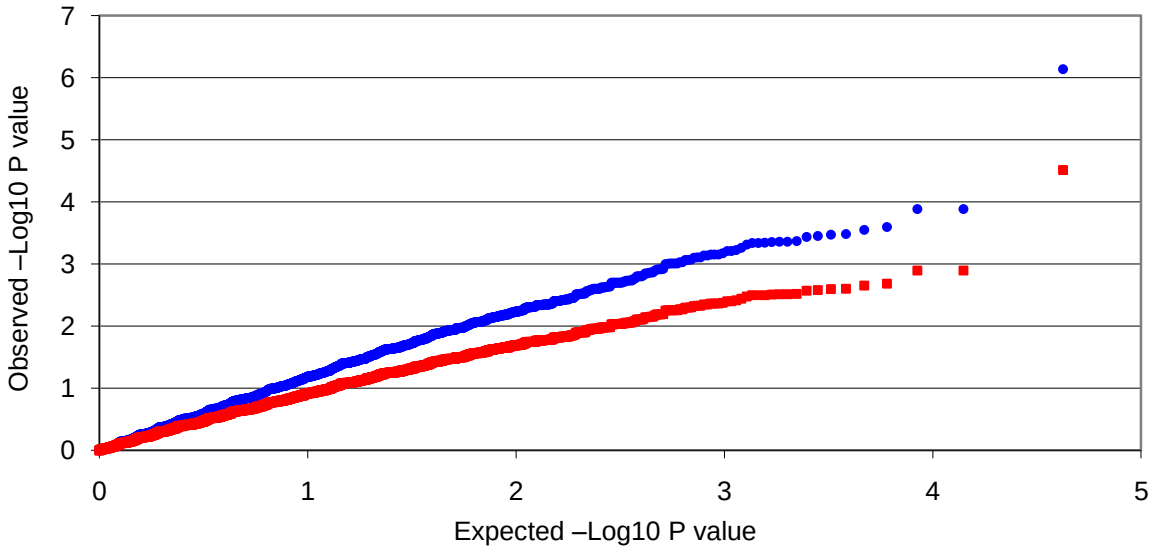


Figure 7.3: A plot showing the difference between the observed and expected results from the GWAS. SNPs are plotted on the x-axis according to the expected –Log10 p value (Red squares) against the observed –Log10 p value (blue circles) on the y-axis. The Genomic control inflation factor is calculated from this plot to represent the difference between observed and expected results. Genomic control inflation factor = 1.41

Figure 7.4 MDS plot of IBS Clusters: Defining Population Structure

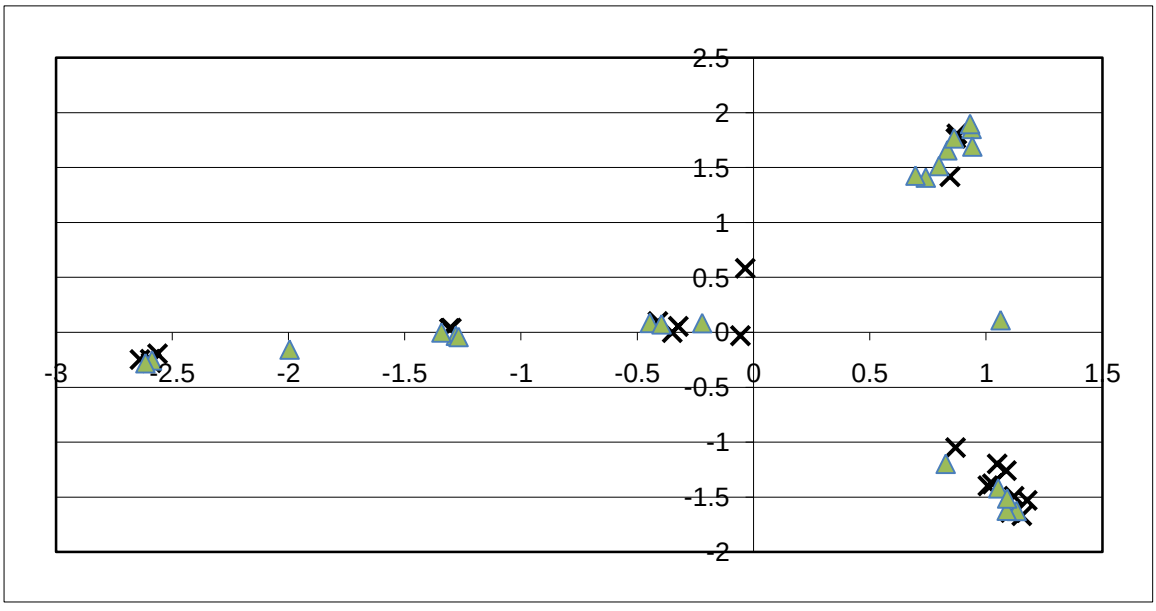


Figure 7.4: A plot showing cases (green triangles) and controls (black crosses) from GWAS clustered by IBS clustering according to the whole genome differences between individuals. Two components are plotted, component one is plotted on the x-axis against component two on the y-axis.

Figure 7.4 represents the population structure of the dogs included on the Illumina canine SNP20. It can be seen that there are two distinct groups and a third more spaced group containing some outliers. The plot also shows that there is no stratification between the cases and controls, suggesting that the three observed population subgroups are not due to differences between cases and controls. It was expected that these subpopulations would represent the two geographical groups included in this study (i.e. USA and Japan). The groups do not, however, cluster according to geographical origin (see Appendix 7). Suggesting that population structure seen here is independent of disease status and geographical location (point of origin); probably reflecting the inbreeding/non-random mating seen in dog breeds. The IBS clusters were used in the meta-analysis of GWAS data to account for population structure. The result this analysis is presented in Figure 7.5.

Figure 7.5 shows the whole genome plot of significance for the "population structure adjusted" results. It can be seen that the significance values are lower than the previous whole genome plot (Figure 7.2), which showed the "unadjusted" results.

Table 7.1 shows the top 40 associated SNPs with cAD from the "population structure unadjusted" and "adjusted" analysis.

Figure 7.5 Population structure adjusted: Whole genome association plot of significance for Canine Atopic Dermatitis

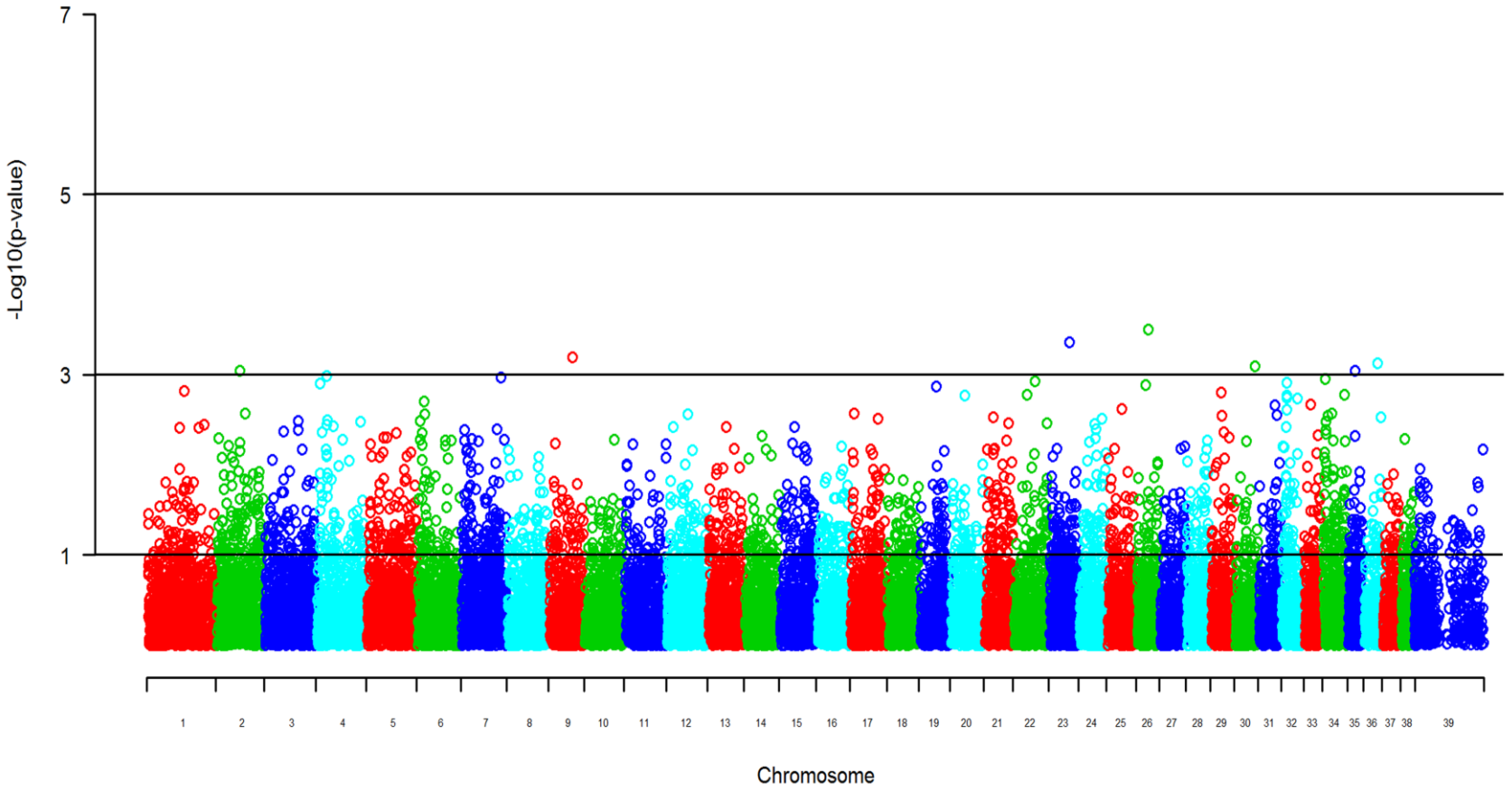


Figure 7.5: A plot showing significance of association of all SNPs from the IBS clustering and meta-analysis ("adjusted" for population structure), with atopic and healthy control phenotypes. SNPs are plotted on the x-axis according to their position on each chromosome against association with atopic dermatitis on the y-axis (shown as -log10 P value).

Table 7.1 Top 40 Associated SNPs from the Population Structure Unadjusted GWAS Results and the Top 40 Associated SNPs from the Population Structure Adjusted GWAS Results

Population Structure: Unadjusted				Population Structure: Adjusted			
SNP	**Chromosome**	**Odds Ratio**	**P value**	**SNP**	**Chromosome**	**Odds Ratio**	**P value**
rs22859255	2	4.4	0.0048	rs21994080	1	0.1	0.0015
rs23545080	3	5.4	0.0019	rs22738169	2	0.2	0.0009
rs23567144	3	0.2	0.0018	rs22808564	2	6.3	0.0027
rs23602938	3	3.8	0.0047	rs24092655	4	0.0	0.0013
rs24872415	3	9.5	$1.27x10^{-06}$	rs24131580	4	45.3	0.001
rs24194054	5	4.8	0.0011	rs24275059	6	40.7	0.0019
rs24267550	5	0.2	0.0042	rs24332727 *	6	0.1	0.0028
rs24318716	6	0.3	0.0076	rs24354997	6	40.7	0.0019
rs24327271	6	0.2	0.0008	rs24462950	7	44.5	0.0011
rs24332727 *	6	0.2	0.0012	rs24556501	9	0.1	0.0006
rs24408651	7	0.2	0.0008	rs24732893 *	12	0.2	0.0028
rs24482628	8	4.1	0.0025	rs22524918	17	NA	0.0027
rs22020166	10	3.8	0.0022	rs22576724	17	0.2	0.0031
rs22114085	10	5.1	0.0008	rs22720467	19	19.8	0.0014
rs22155657	11	5.9	0.0063	rs22864357 *	20	5.6	0.0017
rs22181912	12	0.3	0.0028	rs22991453	21	0.1	0.003
rs22184220	12	4.3	0.0009	rs23029421	22	0.1	0.0017
rs22185491	12	3.4	0.0045	rs23066224 *	22	16.8	0.0012
rs22212359	12	7.2	0.0016	rs23117702	23	0.1	0.0004
rs22212677	12	5.3	0.0049	rs23224877	25	19.5	0.0024
rs24732893 *	12	0.2	0.0007	rs23313105	26	10.5	0.0003

Table 7.1 continued

rs22328353	14	5.8	0.0025	rs23316823	26	0.0	0.0013
rs22523565	16	3.3	0.0067	rs23444049	29	6.3	0.0016
rs8649732	19	0.2	0.0044	rs9132002	29	0.0	0.0029
rs22864357 *	20	4.1	0.0025	rs23622647	30	0.0	0.0008
rs22913552	20	4.6	0.0005	rs23684720	31	5.6	0.0028
rs22915894	20	0.2	0.0017	rs23708599	31	0.2	0.0022
rs8806978	20	3.8	0.0021	rs23767031	32	9.8	0.0012
rs9081246	20	0.2	0.0017	rs23768387 *	32	0.1	0.0019
rs22959786	21	5.9	0.0009	rs23769251	32	0.1	0.0018
rs22989468	21	0.2	0.0057	rs23769439	32	0.1	0.0024
rs23026134	22	5.4	0.0019	rs23770102	32	0.0	0.0017
rs23066224 *	22	7.2	0.0016	BICF2G630250121	33	0.0	0.0022
rs23122280	23	5.4	0.0019	rs23803267	34	0.0	0.0016
rs23428471	28	3.6	0.0036	rs23856926	34	0.2	0.0029
rs23478400	28	3.6	0.0036	rs23865825	34	0.1	0.0027
rs23472497	29	0.2	0.0009	rs23870923	34	0.2	0.0011
rs23690306	31	5.3	0.0003	rs23932942	35	6.3	0.0009
rs23768387 *	32	0.2	0.0031	rs23968934	36	NA	0.003
rs23828846	33	0.2	0.0007	rs23974741	36	NA	0.0007

* Denotes the SNPs identified on population structure adjusted analysis as well as population structure unadjusted analysis

Table 7.2 shows the haplotypes significantly associated with cAD (the individual SNP associations are indicated with an *). A haplotype is significant when the p value for the haplotype is higher than the p value for the individual SNP. Table 7.2 only shows those haplotypes significantly associated with cAD. Figure 7.6 presents results graphically, by indicating the positions of the haplotypes and highlighting which genes are in close proximity. Figure 7.6 (a) shows the haplotypes within the gene RAB3C; (b) shows 4 haplotypes in this area which encompasses 9 different genes on chromosome 6; (c) shows two haplotypes in chromosome 6 which encompass SDK1, ACTB_CANFA and FBXL18; (d) shows 4 overlapping haplotypes most of which fall within RAB7A_CANFA; (e) shows 7 haplotypes on chromosome 20, the majority of which fall within SUCLG2 and LR1G1; and (f) shows 2 haplotypes in intergenic regions on chromosome 20 near 12 genes.

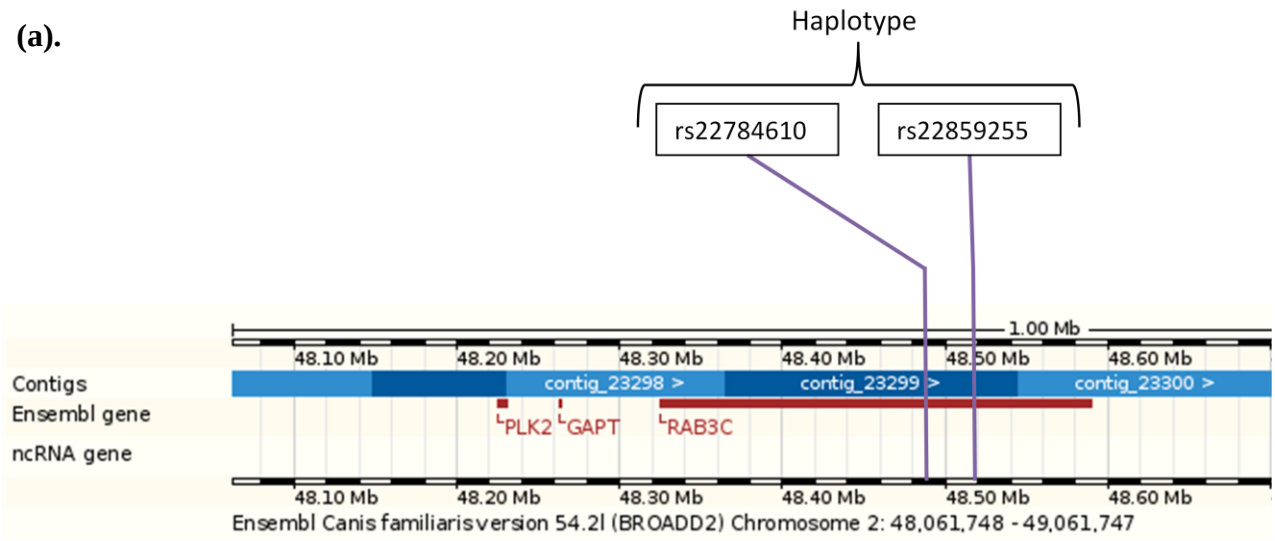
Table 7.2 All Haplotypes significantly associated with Canine Atopic Dermatitis

Chr	Haplotype	P value of haplotype	SNP	P value individual SNP	SNP	P value individual SNP	SNP	P value individual SNP
2	12	0.0004	rs22859255*	0.0048	rs22784610	0.4131	n/a	n/a
6	21	0.0032	rs24327392	1	rs24327060	0.008	n/a	n/a
6	12	0.0044	rs24327060	0.008	rs24259668	0.5711	n/a	n/a
6	22	0.0026	rs24301782	0.0225	rs24354816	0.4683	n/a	n/a
6	22	0.0005	rs24344762	0.4949	rs24332727*	0.0012	n/a	n/a
6	221	0.0002	rs24315819	1	rs24327231	1	rs24327271*	0.0008
6	212	0.0002	rs24327231	1	rs24327271*	0.0008	rs24348619	0.4866
20	111	0.0007	rs22915894*	0.0017	rs8806978*	0.0021	rs22929054	0.405
20	122	0.001	rs9081246*	0.0017	rs22915887	0.1115	rs22915894*	0.0017
20	222	0.001	rs22915887	0.1115	rs22915894*	0.0017	rs8806978*	0.0021
20	221	0.001	rs22915894*	0.0017	rs8806978*	0.0021	rs22929054	0.405
20	21	0.0029	RS8891640	0.0893	rs22892223	0.0093	n/a	n/a
20	11	0.0001	rs22845779	0.2683	rs22913552*	0.0005	n/a	n/a
20	22	0.0007	rs22880046	0.0385	rs22864357*	0.0025	n/a	n/a
20	11	0.0011	rs22883166	0.2626	rs22930928	0.0121	n/a	n/a
20	11	0.0013	rs22930928	0.0121	rs22926113	1	n/a	n/a
20	21	0.0025	rs22913940	1	rs22912434	0.0503	n/a	n/a
20	22	0.004	rs22914527	0.4812	rs22918497	0.0154	n/a	n/a
20	21	0.0047	rs22912434	0.0503	rs22883166	0.2626	n/a	n/a
20	11	0.0049	rs22892983	0.3114	rs22883015	0.0154	n/a	n/a
20	21	0.0099	rs8625504	0.0639	rs22883400	0.35	n/a	n/a

* Denotes that the SNP was significant in the single SNP association analysis

Figure 7.6 Ensembl Chromosome View: A Representation of Haplotype Analysis

(a).



(b).

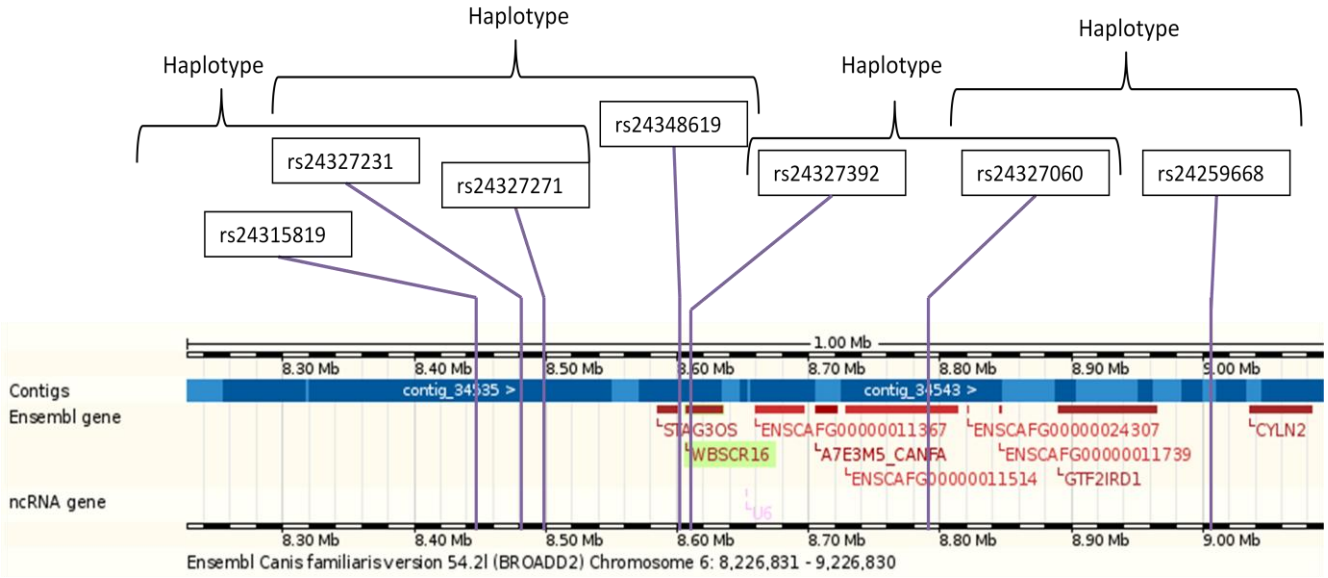
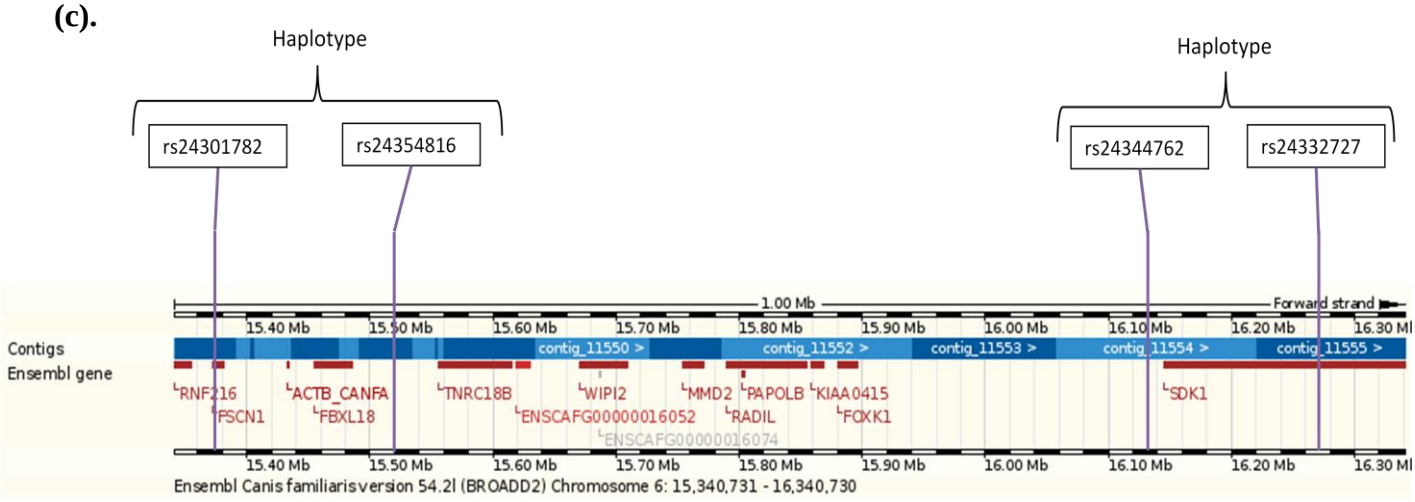


Figure 7.6 continued

(c).



(d).

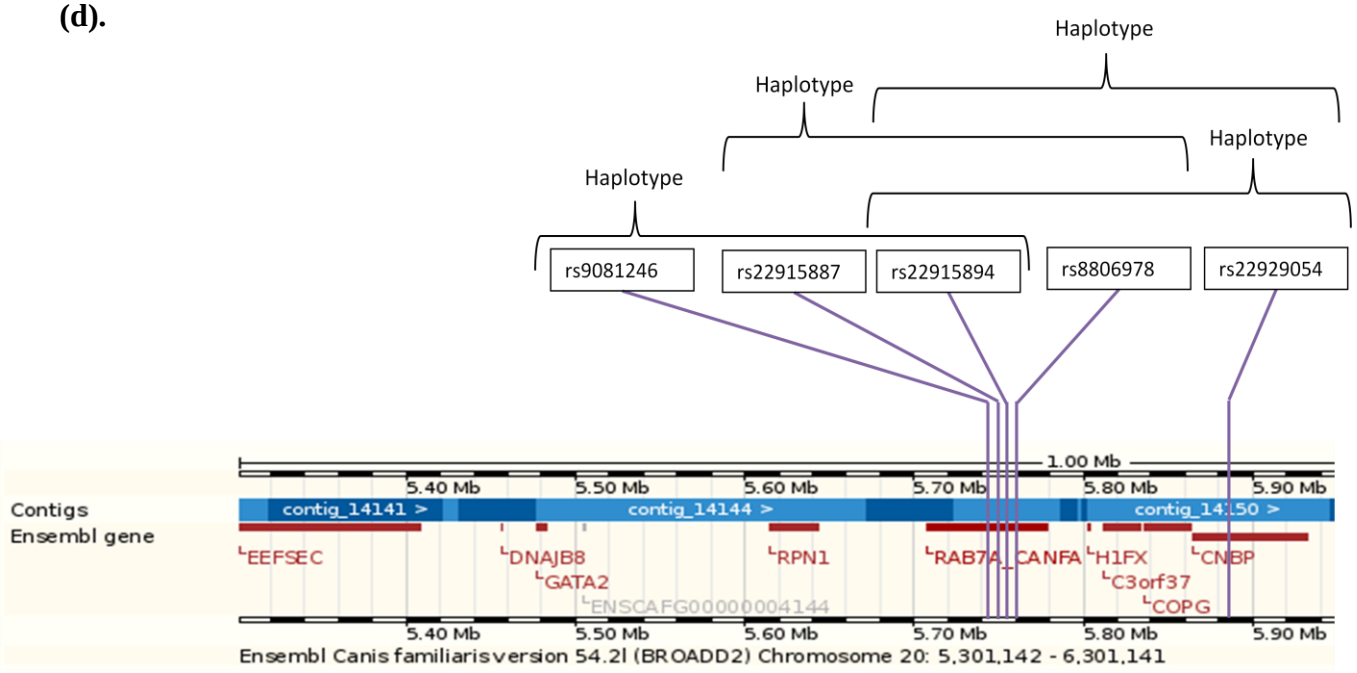


Figure 7.6 continued

(e).

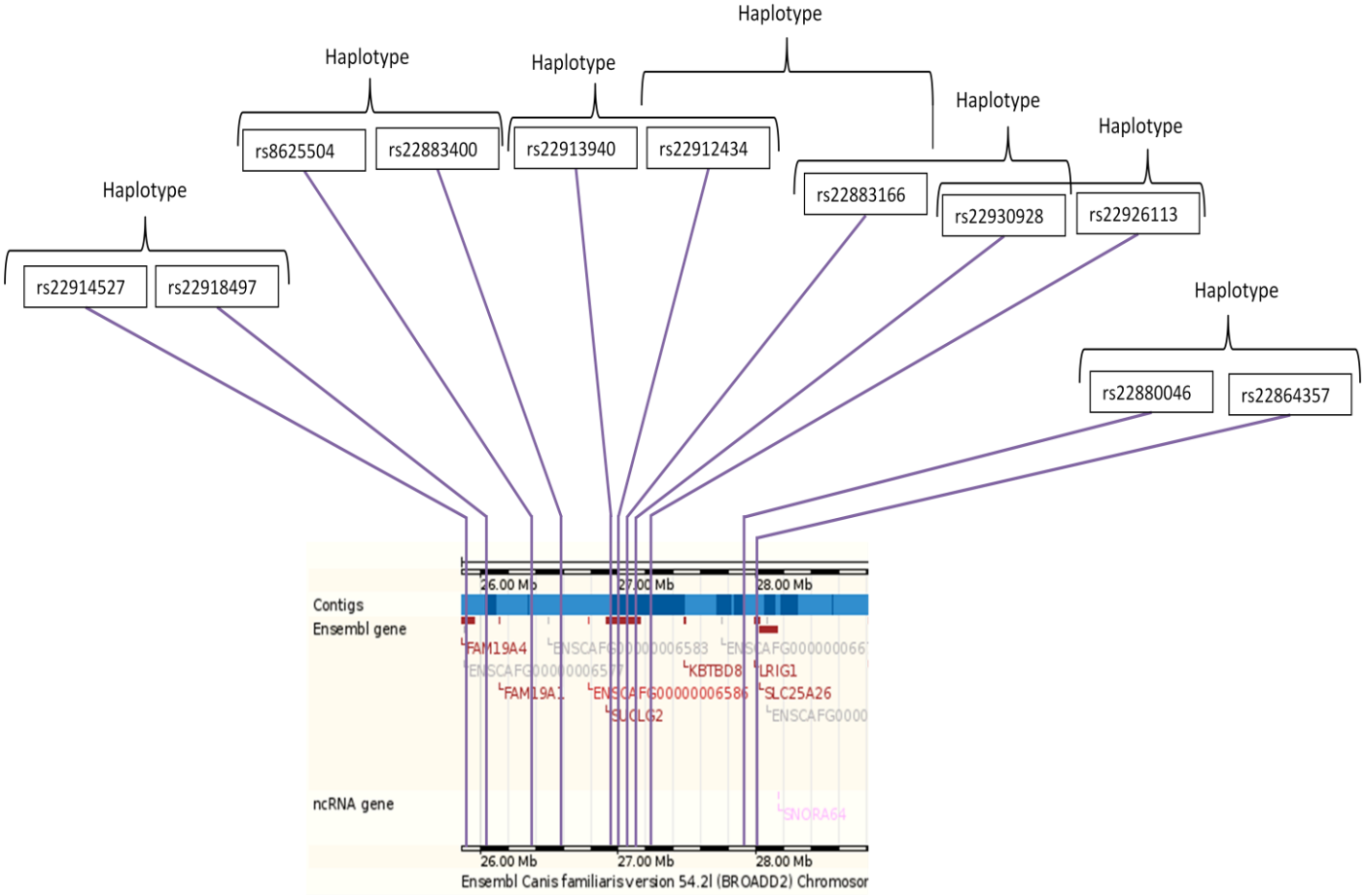


Figure 7.6 continued

(f).

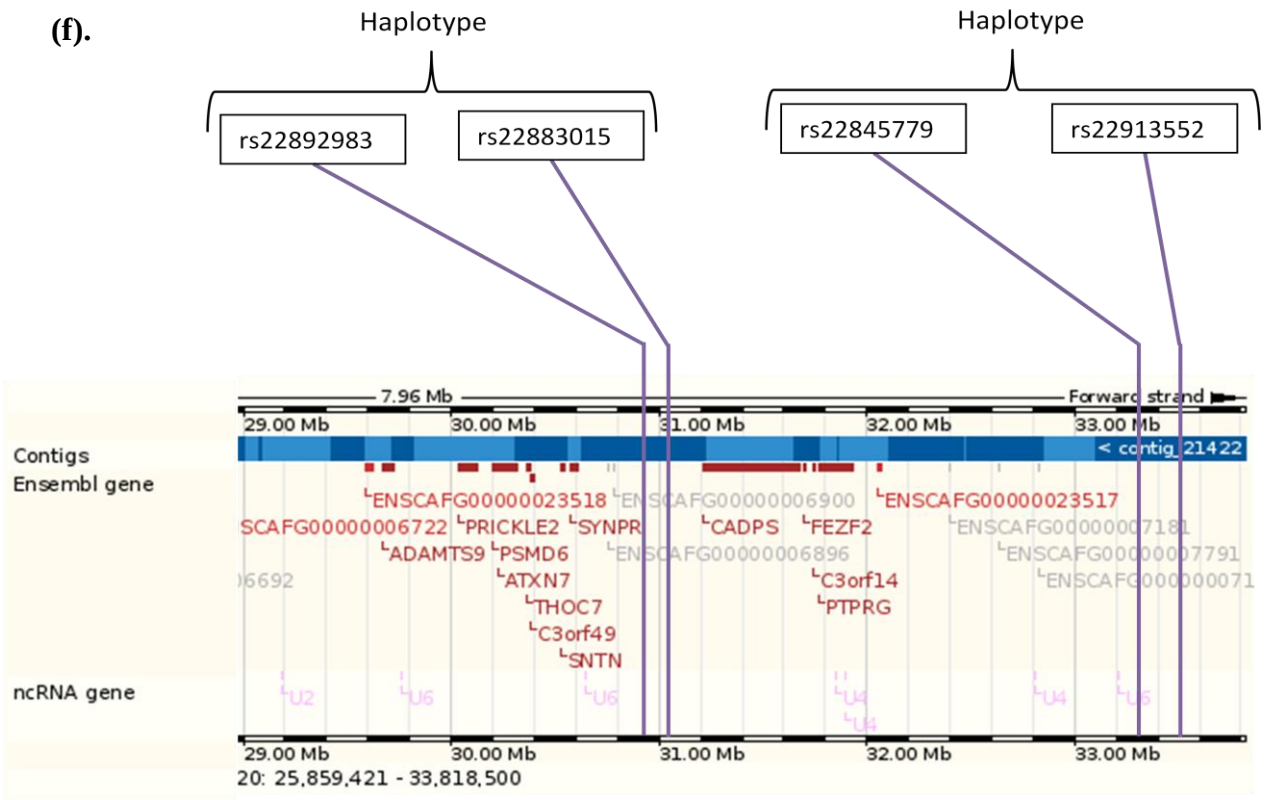


Figure 7.6: Using Ensembl chromosome view, the position of SNPs have been plotted (purple lines), and, using the information in Table 7.2, the significant haplotypes have been indicated in order to give a visual representation of the haplotype data. (a) Chromosome 2, 48061748 to 49061747; (b) Chromosome 6, 8226831 to 9226830 base pairs; (c) Chromosome 6, 15340731 to 16340 730 base pairs; (d) Chromosome 20, 5301142 to 6301141 base pairs; (e) Chromosome 20, 25859421 to 290000000 base pairs; (f) Chromosome 20, 290000000 to 33818500 base pairs.

7.4 Discussion

The results of the GWAS were analysed with and without population structure adjustment. In comparing the top 40 associated SNPs from these two approaches only five SNPs were identified by both tests (marked with* in Table 7.1).

The importance of accounting for population structure in GWAS has been discussed in the context of human[334;340;341], but not in canine studies as this is the first commercially available chip for canine GWAS. Applying the same principles as used in human studies, however, should have accurately defined and accounted for the population structure.

Confounding results due to differences in population structure can mask actual results; especially in complex traits where the effect of each allele is smaller[340]. This may explain higher levels of significance in the "population structure unadjusted" results and only 5 SNPs being identified in both the population structure "adjusted" and "unadjusted" analyses.

The genomic control (GC) inflation factor (calculated by q-q plot, Figure 7.3) can be used to adjust for population structure. Although, it has been shown that GC inflation factor does not accurately correct for population structure[340;341]. This is because GC inflation factors are global and do not account for differences that are specific to each SNP[342;343]. It has been stated that the "potential for stratification based on differences in population structure in case and control groups is different for each SNP"[334], therefore a blanket correction like GC inflation factor is not viable and was not used in this study.

Principle component analysis (PCA) has been shown to account for population structure effectively[334;340;341;343]. However, it has been noted that high LD SNPs affect the population structure findings because they group according to local LD patterns instead of global genome wide population structure. The accuracy of PCA is therefore improved if high LD SNPs are excluded from the analysis[341;344]. In dogs, however, there is insufficient LD

information available to exclude high LD SNPs. As a result this study did not use PCA, and relied on IBS clustering and MDS, which are good alternatives to accurately defining the population structure[334;340]. The issue of high LD SNPs remains.

Population structure results (Figure7.4) and the Q-Q plot (Figure 7.3) shows differences within the golden retriever breed. It was expected that these differences would be due to different points of origin as these samples were obtained from veterinary practices in Japan and the US. The population structure, however, did not group according to point of origin and it seems there is a discreet population structure present. Unknown familial relationships between cases and controls can result in them appearing to have their own population structure[334]. This has implications for population structure in dogs because they are inbred, especially amongst pedigrees that have a limited number of or particularly popular sires[345]. No familial relationship information for the golden retrievers used on the array was available, therefore it is possible that the results grouped according to a unknown familial relationship.

The use of family based cases and controls excludes false positives due to population structure, and this principle has been effectively applied to human GWAS[346;347]. This could be a viable approach for future canine GWAS studies because of the population structure issues. However, first degree relatives are required for analysis and, as discussed in Chapter 1, obtaining enough samples for family studies is difficult.

The "population structure adjusted" results may be more accurate because there are such large differences within the breed they need to be accounted for. However, the methods used rely on a dense set of trait independent and unlinked SNPs[340], and, given the lack of LD information on dogs, it is impossible to say whether this has been accounted for. It cannot be concluded that the "population structure adjusted" results are the more accurate results and therefore both sets of results need to be considered.

Three of the 5 SNPs found (on both "population structure adjusted" and "unadjusted" analyses, and the majority of the SNPs identified on either analysis) were present in intergenic regions. Further fine mapping and re-sequencing will be required if they are found to be significant in the subsequent validation study. Only a few of the potentially associated SNPs were found in or around genes.

SDK1 and LRIG1 SNPs were identified by both analyses. SDK1 is cell adhesion protein containing immunoglobulin-like and fibronectin domains. It is considered to only operate in axonal terminals and therefore SDK1 has no apparent direct functional link to cAD. LRIG1, on the other hand, acts as a negative regulator of signalling through enhancement of receptor ubiquitination (the process of protein inactivation by addition ubiquitin, to allow transport of the protein to the proteasome for degradation) and accelerated intracellular degradation, suggesting a possible function in cAD. Interestingly haplotype analysis demonstrated that both SDK1 (Figure 7.6c) and LRIG1 (Figure 7.6f) have more significant p values when analysed with certain intergenic SNPs (see Table 7.2), suggesting a haplotypic effect.

In the "population structure unadjusted" results SNPs within RAB3C, SORCS2, PROM1, FXYD8, KHDRBS2, TFEB, COL19A1, RAB7A, RARB, MAPK8, SEMA5B and ENSCAFG00000004115 were identified (Figure 7.2, Table 7.1).

RAB3C has a potentially relevant function to the pathogenesis of AD as it codes for a protein transporter found on lipid anchors[348]. The importance of lipids and the skin barrier has been demonstrated multiple times in both human and canine AD[3;4;45;349;350]. The associated SNPs (rs22859255, rs22784610) show a more significant p value when analysed together, suggesting that they form a haplotype which is significantly associated with cAD (see Table 7.2, Figure 7.6a).

COL19A1 is a collagen gene with roles in the formation and maintenance of the extracellular matrix[351], changes to which may affect the pathogenesis of AD. TFEB recognizes E-box sequences in the heavy-chain immunoglobulin enhancer[352]; this could have implications for AD due to the potential effects on the recognition of allergens.

RAB7A produces a protein which is a cellular component of the melanosome, containing melanin for skin pigmentation. The association with cutaneous metabolism suggests that RAB7A could be involved in cAD although the precise pathogenesis is unclear. Multiple significant haplotypes with RAB7A intronic and downstream SNPs have been demonstrated (see Table 7.2, Figure 7.6d), suggesting that further fine mapping of the region could be beneficial.

MAPK8 (JNK1) is involved in the JNK signalling pathway[353] which activates in response to environmental stress and pro-inflammatory cytokines[354]. In T-cells, JNK1 and JNK2 are required for polarized differentiation of T-helper cells into Th1 cells. These functions are therefore potentially relevant to the immunology of AD.

The IBS cluster/meta-analysis ("population structure adjusted") results found SEZ6L, RAB11-FIP4, precursor C5orf46, Solute carrier family 35 member F3, PREX2, F-box only protein 41 ENSCAFG00000004544 (novel gene), ENSCAFG00000007979 (novel gene) and ENSCAFG00000015477 (novel gene) to have an association with cAD. However, none of these genes show immediately obvious functions which may be involved with the pathogenesis of cAD (Figure 7.5 and Table 7.2).

The haplotype analysis showed increased significance in two intergenic SNPs (rs24327271 [Figure 7.6b] and rs22913552 [Figure 7.6f]) when analysed with other nearby intergenic SNPs, suggesting that further haplotype analysis and fine mapping of these regions is required. There were also other haplotype associations found with SNPs that were not

significant on the single SNP analysis. For example, SNPs in SUCLG2 (Succinyl-CoA ligase [GDP-forming] subunit beta, mitochondrial) which is involved in carbohydrate metabolism[355] and FAM19A4, a brain specific protein of unknown function[356] formed separate significant haplotypes (Figure 7.6d). Moreover SNPs within Williams-Beuren syndrome chromosomal region 16 protein (WBSCR16), ENSCAFG00000011514 (an orthologue of Williams-Beuren syndrome chromosomal region 6 protein [WBSCR6]), FSCN1 and STAG3OS also formed significant haplotypes with intergenic SNPs (Figure 7.6b).

WBSCR6 produces a protein which is necessary for induction of immunoglobulin heavy-chain transcription[357] and FSCN1 organizes filamentous actin into bundles and associates with beta catenin[358]. These two functions could be linked with AD pathogenesis, although further fine mapping and analyses of these areas is required before conclusions can be drawn.

Haplotype analysis provides a useful way to gain more information from GWAS data. Because of the high LD in dogs, it is difficult to know whether the haplotypes identified are causal or simply in linkage with other causative alleles. Until a canine HAPMAP equivalent is available it will be difficult to further differentiate or investigate this fully.

In conclusion, using GWAS identified novel candidate SNPs in cAD for a large-scale case-control comparison/validation study. However, the numbers of dogs were low, increasing the risk of detecting false positive associations. This was statistically corrected for, although this does not eliminate false positives all together. It would have been better to use a large number of cases and controls to enhance the statistical power and confidence in the results. The issue of population structure indicates that in future studies sample selection for GWAS should use ancestry informative markers to initially define population structure before selecting grouped individuals for inclusion on GWAS[334]. Although this would reduce the already small sample sizes available. Finally, it is important to note that while the canine

SNP20 array is defined as having good coverage across the genome, the success of the study depends on whether the array includes the causative allele. In this array there are large gaps in coding regions, and Filaggrin, for example, which appears to be of great importance in hAD[228;230;231;359;360], is not covered by the Illumina canine SNP20.

These results therefore need to be validated and should be considered a first sweep or novel SNP identification method. By performing further genotyping validation it is hoped that true associations with the disease phenotype will be found.

Chapter 8

GWAS Validation Study

8. GWAS Validation Study

This work is presented in: **Wood, S.H.**, Ke, X., Nuttall, T., McEwan, N., Ollier, W.E., Carter, S.D., (2009) Genome Wide Association Analysis of Canine Atopic Dermatitis and Identification of Disease Related SNPs, *Immunogenetics*, 61(11-12):765-72

8.1 Introduction

In Chapter 7, SNPs associated with cAD were identified by GWAS. Validation of these results is necessary to have confidence in the associations found, as a small number of dogs were included on the GWAS array, which could lead to false positive associations. To validate the GWAS results, the associated SNPs were genotyped by another method.

Sequenom is a high throughput genotyping system that uses multiplex PCR followed by a single base primer extension reaction. Subsequent analysis by MALDI-TOF spectrophometry ensures accurate identification of the alleles present. As Sequenom is a multiplex system, it is relatively inexpensive and only requires small quantities of DNA sample.

In Chapter 7 the GWAS data were analysed in two ways: Firstly using the Fisher's exact test, which does **not** take into account population structure present in the dogs studied and could be prone to false positive associations[341], and secondly, a meta-analysis, which using IBS clustering and MDS, defines and accounts for population structure[334]. When comparing the results from these two methods only five SNPs were identified in both analyses as potentially associated with cAD.

The disparity between the two analyses was not unexpected, as the results observed can be masked by differences in population structure, especially in complex traits where the effect of each allele is small[340]. As the results were different and IBS clustering/MDS has not

been used to identify population structure in dogs before, the decision was taken to validate the top 40 results from the analysis using the Fisher's exact test ("population structure unadjusted") and the top 40 SNPs from IBS and MDS analysis ("population structure adjusted"). The five SNPs identified by both analyses were used as quality control markers of inter-assay variation.

By validating results from both "population structure adjusted" and "unadjusted" analyses it was hoped that we would be able to identify whether IBS clustering was adequately identifying population structure. Theoretically, if IBS/MDS was identifying true associations and Fisher's exact test was detecting false positives due to population structure, the IBS/MDS "population structure adjusted" analysis should show more positive associations when genotyped by Sequenom, than the "population structure unadjusted" analysis. The validation and comparison of these two analyses could have implications in future canine GWAS study design and analysis. Moreover, if positive associations were found after validation it would indicate that the use of small sample size GWAS in canines for novel SNP identification is of scientific merit.

The study design is presented in Figure 8.1, which outlines the selection of SNPs from the GWAS and the subsequent analyses that were performed on the Sequenom validation results in this chapter.

Figure 8.1 GWAS Validation Study Design

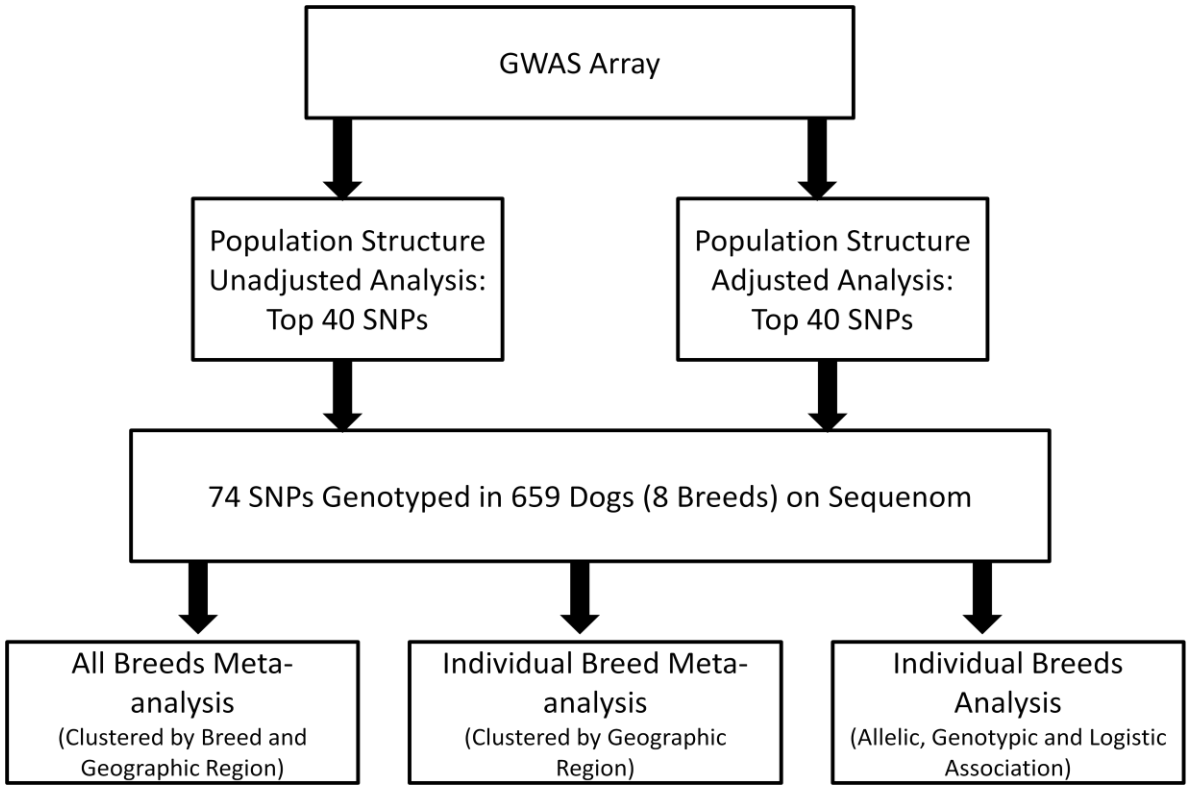


Figure 8.1: GWAS array was performed in Chapter 7 and the results were analysed in two ways; "population structure adjusted" and "population structure unadjusted". The top 40 from both these analyses were typed on the Sequenom to validate the results. The analysis was three fold; meta-analysis on all breeds, individual breed meta-analysis and individual breeds split by geographic region association tests.

The validation of the GWAS data was carried out in 659 dogs across 8 breeds, including the golden retriever as used in the GWAS (Chapter 7). The inclusion of multiple breeds complicates the analysis of the Sequenom data. Firstly, differences in breed and geographical region need to be taken into account, requiring meta-analysis of all 8 breeds clustered by breed and geographical region. Secondly, there is no evidence that cAD will be caused by the same genetic variations in different breeds therefore an individual breed analysis was required. This was done in two ways: meta-analysis on individual breeds clustered by

geographic region, e.g. all golden retrievers; and allelic association, genotypic model and logistic association tests performed on individual breeds from specific geographical regions, e.g. three separate analyses on British, US and Japanese golden retrievers. Analysing the data in this way allowed population structure caused by geographical region and inter-breed differences to be accounted for.

The aim of this study was to validate results obtained from a GWAS in a larger sample set using the Sequenom SNP genotyping platform. Validation of the GWAS results that show SNP associations with cAD will help validate the use of small sample size GWAS as a novel SNP discovery method for potentially causative SNPs in cAD.

8.2 Methods

Atopic dogs and healthy controls – breed, sex and age matching

Disease diagnosis, sample collection and processing of samples from dogs with and without cAD were performed as detailed in Chapter 2. Five hundred and four dog DNA samples were breed, age and sex matched. One hundred and fifty five samples were only breed matched. In total, 659 DNA samples were included on the Sequenom array, 41 of which were previously included on the GWAS array. Eight breeds were included in this study. Table 8.1 summerises the breakdown of breed, geographical region and disease status.

Table 8.1 Samples included in Sequenom genotyping

	CASE				CONTROL				
Breed	UK	USA	JAPAN	Total	UK	USA	JAPAN	Total	**Grand Total**
Boxer	11	-	-	11	24	-	-	24	**35**
GSD	10	9	-	19	23	15	-	38	**57**
Labrador	23	33	8	64	75	43	11	129	**193**
G. Retriever	4	23	13	40	28	21	17	66	**106**
Shiba Inu	-	-	23	23	-	-	33	33	**56**
Shih Tzu	2	6	39	47	12	2	48	62	**109**
Pit Bull	-	20	-	20	-	17	-	17	**37**
WHWT	14	4	-	18	34	14	-	48	**66**
Totals				**242**				**417**	**659**

Primer and probe design

Table 8.2 summerises the SNPs selected from the GWAS in Chapter 7. Full primer and probe sequences for these SNPs are listed in Appendix 8. Primers and probes could not be adequately designed for BICF2G630250121, and 5 SNPs were identified in both analyses, bringing the total number of SNPs genotyped to 74.

Table 8.2 Top 40 SNPs identified on GWAS from Population Structure Unadjusted Analysis and Population Structure Adjusted Analysis

Population Structure: Unadjusted				Population Structure: Adjusted			
SNP	**Chromosome**	**Odds Ratio**	**P value**	**SNP**	**Chromosome**	**Odds Ratio**	**P value**
rs22859255	2	4.4	0.0048	rs21994080	1	0.1	0.0015
rs23545080	3	5.4	0.0019	rs22738169	2	0.2	0.0009
rs23567144	3	0.2	0.0018	rs22808564	2	6.3	0.0027
rs23602938	3	3.8	0.0047	rs24092655	4	0.0	0.0013
rs24872415	3	9.5	$1.27x10^{-06}$	rs24131580	4	45.3	0.001
rs24194054	5	4.8	0.0011	rs24275059	6	40.7	0.0019
rs24267550	5	0.2	0.0042	rs24332727 *	6	0.1	0.0028
rs24318716	6	0.3	0.0076	rs24354997	6	40.7	0.0019
rs24327271	6	0.2	0.0008	rs24462950	7	44.5	0.0011
rs24332727 *	6	0.2	0.0012	rs24556501	9	0.1	0.0006
rs24408651	7	0.2	0.0008	rs24732893 *	12	0.2	0.0028
rs24482628	8	4.1	0.0025	rs22524918	17	NA	0.0027
rs22020166	10	3.8	0.0022	rs22576724	17	0.2	0.0031
rs22114085	10	5.1	0.0008	rs22720467	19	19.8	0.0014
rs22155657	11	5.9	0.0063	rs22864357 *	20	5.6	0.0017
rs22181912	12	0.3	0.0028	rs22991453	21	0.1	0.003
rs22184220	12	4.3	0.0009	rs23029421	22	0.1	0.0017
rs22185491	12	3.4	0.0045	rs23066224 *	22	16.8	0.0012
rs22212359	12	7.2	0.0016	rs23117702	23	0.1	0.0004
rs22212677	12	5.3	0.0049	rs23224877	25	19.5	0.0024
rs24732893 *	12	0.2	0.0007	rs23313105	26	10.5	0.0003

Table 8.2 continued

rs22328353	14	5.8	0.0025	rs23316823	26	0.0	0.0013
rs22523565	16	3.3	0.0067	rs23444049	29	6.3	0.0016
rs8649732	19	0.2	0.0044	rs9132002	29	0.0	0.0029
rs22864357 *	20	4.1	0.0025	rs23622647	30	0.0	0.0008
rs22913552	20	4.6	0.0005	rs23684720	31	5.6	0.0028
rs22915894	20	0.2	0.0017	rs23708599	31	0.2	0.0022
rs8806978	20	3.8	0.0021	rs23767031	32	9.8	0.0012
rs9081246	20	0.2	0.0017	rs23768387 *	32	0.1	0.0019
rs22959786	21	5.9	0.0009	rs23769251	32	0.1	0.0018
rs22989468	21	0.2	0.0057	rs23769439	32	0.1	0.0024
rs23026134	22	5.4	0.0019	rs23770102	32	0.0	0.0017
rs23066224 *	22	7.2	0.0016	BICF2G630250121	33	0.0	0.0022
rs23122280	23	5.4	0.0019	rs23803267	34	0.0	0.0016
rs23428471	28	3.6	0.0036	rs23856926	34	0.2	0.0029
rs23478400	28	3.6	0.0036	rs23865825	34	0.1	0.0027
rs23472497	29	0.2	0.0009	rs23870923	34	0.2	0.0011
rs23690306	31	5.3	0.0003	rs23932942	35	6.3	0.0009
rs23768387 *	32	0.2	0.0031	rs23968934	36	NA	0.003
rs23828846	33	0.2	0.0007	rs23974741	36	NA	0.0007

* Denotes the SNPs identified on population structure adjusted analysis as well as population structure unadjusted analysis

Quality control, plex dilutions and plate setup

See Chapter 2 for details of quality control, plexing and plate setup for the Sequenom.

PCR and SAP treatment

The PCR assays were all performed in a 384-well plate format. A 5µl reaction volume was used per well. This consisted of 0.1µl Hot start Taq (5U/µl, Qiagen, Crawley, UK), 0.5µl each of 500 nM forward and reverse primers (Metabion, Germany), 0.325µl $MgCl_2$ (25mM, Qiagen, Crawley, UK), 0.1µl DNTPs (25mM), 0.625µl x10 buffer (Qiagen, Crawley, UK) 0.35µl distilled water and 2.5µl of DNA at 10ng/µl. The amplification was performed as follows: 94ºC for 15 minutes, then 94ºC for 20 seconds, 56ºC for 30 seconds and 72ºC for 1 minute (35 cycles), and finally 72ºC for 3 minutes.

Shrimp Alkaline Phosphatase (SAP) clean was performed by adding 1.53µl distilled water, 0.17µl 10x SAP buffer (Sequenom, San Diego, CA, USA) and 0.3µl SAP enzyme (1U/µl) to each well, to make a total reaction volume of 7µl. This was incubated at 37ºC for 40 minutes and 85ºC for 5 minutes.

iPLEX reaction and Resin clean

The iPLEX reactions were all performed in a 384-well plate format. To the SAP cleaned plate, 0.619µl distilled water, 0.2µl x10 iPLEX buffer (Sequenom, San Diego, CA, USA), 0.2µl iPLEX termination mix (Sequenom, San Diego, CA, USA), 0.041µl iPLEX enzyme (Sequenom, San Diego, CA, USA) and 0.94µl probe mix (7µM and 14µM, Sequenom, San Diego, CA, USA) were added to make a total reaction volume of 9µl. The cycling conditions were as follows: 94ºC for 30 seconds, then 94ºC for 5 seconds, 52ºC for 5 seconds and 80ºC for 5 seconds (40 cycles), and finally 72ºC for 3 minutes.

A resin clean was performed by adding 16µl of distilled water per well and the appropriate amount of resin (Clean resin, Sequenom, San Diego, CA, USA) per well (3mg if the plex was lower than 5 and 6mg if the plex was above 6). The plate was inverted for 10 minutes allowing the resin to settle each time. The plate was then centrifuged at 4000rpm for 15 minutes.

Spotting and running arrays

The sample plates and arrays (Spectro chip®, Sequenom, San Diego, CA, USA) were spotted using Massarray™ Nanodispenser. Sequenom Samsung 3pt calibrant (Sequenom, San Diego, CA, USA) was also spotted to allow the MALDI-TOF spectrophotometer (Sequenom, San Diego, CA, USA) to take a control reading. Spots were read by the MALDI-TOF spectrophotometer (Sequenom, San Diego, CA, USA).

Data collection, calling and analysis

The Spectro-aquire and mass array typer software packages (Sequenom, San Diego, CA, USA) interpreted these data and automatically called the genotypes from the spectra. These data were reviewed using Typer analyser V3.4.0.18 (Sequenom, San Diego, CA, USA) and the data were exported into Excel format for statistical analysis.

Statistical analysis

PLINK[338] was used for all the downstream analysis. All data were subject to quality control procedures; only SNPs which conformed to Hardy-Weinberg expectations, exhibited 90% genotyping success and had a minor allele frequency ≥ 0.01were included in subsequent analysis. On this basis eleven of the 74 SNPs tested were excluded (rs23567144, rs22020166, rs22155657, rs9081246, rs22959786, rs23122280, rs21994080, rs23444049, rs23803267, rs23856926 and rs23029421). A variety of association tests were undertaken (see Figure 8.1

for overview). Using PLINK, each breed was analysed separately according to geographic region, and allelic association, logistic association and genotypic model tests were performed. Meta-analysis (Cochran-Mantel-Haenszel test for 2x2xK stratified tables) was used in two ways. Firstly to analyse all 8 breeds together taking into account breed and geographic region (defined by user from supplied phenotype information). Secondly, meta-analysis was used on individual breeds where samples were obtained from different geographic regions, e.g. golden retrievers were obtained from USA, UK and Japan. It is considered that this approach eliminates false positives due to population substructure[334]. P values were corrected for multiple testing by using the number of independent loci (38 in this study).

8.3 Results

The 41 samples included in both the GWAS discovery study (Chapter 7) and Sequenom arrays showed an average concordance of 99% in all genotyped SNPs. Of the 5 SNPs identified in both analyses, 3 were genotyped on each Sequenom array to act as an inter-array control revealing 100% concordance between the arrays.

Sequenom SNP data from all breeds were combined and analysed using meta-analysis to account for the differences between breeds and geographic region. Table 8.3 shows the corrected p value significant SNPs from the meta-analysis for all 8 breeds, and compares the results to the GWAS (Chapter 7) and a combined meta-analysis of the GWAS data and the Sequenom data. A complete summary of the results from the meta-analysis are presented in Appendix 9. A conditional regression on the meta-analysis data from Table 8.3 is presented in Table 8.4.

Table 8.5 shows the results from the individual breed analysis as a matrix of corrected p values obtained from the allelic association tests and individual breed meta-analysis. Where indicated, the associations were significant on the logistic and genotypic association tests.

The complete results from all of the statistical tests undertaken on individual breeds are presented in Appendices 10 and 11.

Seventy four SNPs were typed on the Sequenom platform, 11 of which were excluded due to low genotyping success. Out of the 63 remaining SNPs, 30 had statistically significant corrected p values (summarised in Table 8.5). The meta-analysis on all 8 breeds showed strong association with cAD in 2 SNPs (Table 8.3).

Table 8.3 Comparison of the Validation Study Meta-Analysis, GWAS Analysis and a Combined Analysis of Validation and GWAS Data

SNP	**GWAS Discovery Study**		**Validation Study - Meta analysis**		**Combined Validation study & GWAS meta-analysis**	
rs number	*Odds Ratio*	*Corrected P value*	*Odds ratio*	*Corrected P value*	*Odds Ratio*	*Corrected P value*
RS22114085	5.1	0.00084	2.0	0.00014	2.0	$4.8x10^{-05}$
RS23472497	0.2	0.00086	0.6	0.0015	0.6	0.0011

Table 8.4 Conditional Regression of Significant Meta-analysis SNPs

	ADD	**RS22114085**
RS23472497	0.00056	0.00399
RS22114085	0.00399	0.00056

Table 8.4 shows the results of a conditional regression on the two SNPs identified by meta-analysis as statistically significant in cAD. The p values were significant in this test indicating that the two SNPs are independent of each other.

Table 8.5 Individual Breed Analysis: Matrix of Corrected P value Significant SNPs

	Boxer		**GSD**			**Labrador**		**Golden retriever**			**Shiba Inu**	**Shih Tzu**	**Pit Bull**	**WHWT**		
Rs number	*UK*	*All*	*UK*	*USA*	*All*	*USA*	*UK*	*All*	*USA*	*UK*	*Japan*	*Japan*	*USA*	*All*	*UK*	*USA*
RS22114085	-	-	-	-	-	-	$2.0x10^{-4}$**	-	-	-	-	-	-	-	-	-
RS22184220	-	-	-	0.034	-	-	-	-	-	-	-	-	-	-	-	-
RS22523565	-	-	-	-	-	-	-	-	0.022	-	-	-	-	-	-	-
RS22859255	-	-	-	-	-	-	-	0.001	-	-	-	-	-	-	-	-
RS22915894	-	-	-	-	-	-	0.002**	-	-	$1.6x10^{-6}$*	-	-	-	0.033	-	0.006*
RS23026134	-	-	-	-	-	0.009*	-	-	-	0.018*	-	-	-	-	-	-
rs23472497	-	-	-	-	-	-	-	0.032	-	-	-	-	-	-	-	-
RS23545080	-	-	-	-	-	-	-	0.045	-	$1.0x10^{-6}$*	-	-	-	-	-	-
RS23602938	-	-	-	-	-	-	-	0.018	-	-	-	-	-	-	-	-
RS23690306	-	-	-	-	-	-	-	-	-	0.024*	-	-	-	-	-	-
RS24194054	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.032*	-
RS24327271	-	-	-	-	-	-	-	0.01	0.014	-	-	-	-	-	-	-
RS24408651	-	-	-	-	-	-	-	0.019	-	-	-	-	-	-	-	-
RS24482628	-	-	-	-	-	-	-	-	-	-	-	-	-	0.023	0.006*	-
RS24872415	-	-	0.009	-	-	-	-	-	-	-	-	-	-	-	-	-
RS8806978	-	-	-	-	-	-	0.006*	-	-	$2.3x10^{-6}$*	-	-	-	0.033	-	0.006*
RS22720467	-	-	0.001	-	-	-	0.032	-	-	-	-	-	-	-	-	-
RS23224877	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.029	-
RS23316823	-	-	-	-	-	-	-	-	0.007*	-	-	-	-	-	-	-
RS23708599	-	-	-	-	-	-	-	-	0.013	-	-	-	-	-	-	-
RS23767031	-	-	-	-	-	-	-	-	-	-	$8.0x10^{-4}$**	-	-	-	-	-
RS23769251	-	-	0.006	-	-	-	-	-	-	-	-	-	-	-	-	-
RS23770102	-	-	-	-	-	-	-	-	0.027	-	-	-	-	-	-	-

Table 8.5 continued

RS23968934	-	-	-	-	-	-	-	-	-	$4.0x10^{-4}$	-	-	-	-	-	-
RS24092655	-	-	-	-	-	-	-	-	-	-	0.004	-	-	-	-	-
RS24275059	-	-	0.047	-	-	-	0.035	-	-	-	-	-	-	-	-	-
RS24354997	-	-	0.047	-	-	-	-	-	-	-	-	-	-	-	-	-
RS9132002	-	-	-	-	-	-	0.002*	-	-	-	-	-	-	-	-	-
RS22864357	-	-	-	-	-	-	-	-	-	$2.0x10^{-4}$*	-	-	-	-	-	-
RS24332727	-	-	-	-	-	-	-	-	-	0.007	-	-	-	-	-	-

* denotes that this association is significant on trend analysis using the genotypic model. ** denotes that this association is significant on trend analysis using the genotypic model and logistic association using the additive model. For the results of all statistical tests undertaken on individual breeds, which include p values, upper and lower confidence intervals, and odds ratios (see appendices 10 and 11).

8.4 Discussion

Meta-analysis of the eight breeds under study found two independent associations, which, represent a cAD susceptible SNP (RS22114085) and a cAD protective SNP (RS23472497). These two SNPs are both intergenic within chromosomes 10 and 29 respectively. Both of these SNPs were selected from the "population structure unadjusted" GWAS analysis. The breeds were also analysed separately for associations, revealing that rs22114085 was significantly cAD-associated in UK Labradors (see Table 8.5). All golden retrievers showed a significant cAD association with rs23472497, but individual breed associations for these two SNPs were not seen in the other breeds. This could have been because the small sample sizesanalysed affected statistical power. These results suggest that the associated SNPs, rs22114085 and rs23472497, may represent novel loci suitable for further re-sequencing and fine mapping in cAD research.

An individual breed analyses was also undertaken, as it was unknown whether cAD has a common genetic cause in all breeds. The results of this analysis are presented in Table 8.5. Thirty SNPs were associated with cAD in various breeds; the majority of the associations found were in intergenic SNPs (SNPs not within genes). Only the associations with SNPs within genes will be discussed here as the role of SNPs in intergenic regions is unknown.

In all golden retrievers, two intronic SNPs, rs23602938 (PROM1) and rs22859255 (RAB3C), were found to be significantly associated with cAD. RAB3C has a potentially relevant function contributing to the pathogenesis of cAD as it codes for a protein transporter found on lipid anchors[348], and the importance of lipids and the skin barrier has been previously demonstrated in both human and canine AD[3;4;44;349;350]. Moreover, using the GWAS data I demonstrated that the cAD-associated SNP forms a significant haplotype with another intronic SNP from RAB3C (rs22859255 and rs22784610, p value=0.000386, see

Chapter 7). This, coupled with the validation of this SNP by Sequenom, indicates that the association is of potential importance in cAD. However, rs22784610 was not genotyped in this validation study because it was not individually significant on the GWAS.

In Labradors meta-analysis revealed no significant associations, but when the analysis was split into geographic regions, UK Labradors showed associations between cAD and an intronic SNP within RAB7A, rs22915894, and rs8806978, which is downstream of RAB7A. These associations were also seen in WHWTs. RAB7A, produces a protein which is a cellular component of the melanosome, containing melanin for skin pigmentation. As cAD is a dermatological disease, the function of RAB7A could be relevant to the pathogenesis. Identification of an association with these two intronic SNPs may also suggest that there is a haplotypic effect; in Chapter 7 a significant haplotype between these two SNPs was demonstrated (p=0.0007) from the GWAS data. The same analysis cannot be performed on this data, however, because the data are not genome wide or of sufficient density. Further investigation by fine mapping/re-sequencing is therefore required to confirm this haplotype. These SNPs were also significant before multiple testing corrections in all breeds on the validation study meta-analysis (see Appendix 7). This could suggest that SNPs in linkage in this region could be causative or have a haplotypic effect in other breeds.

RS24872415, an intronic SNP within SORCS2, which was the most significant SNP on the "population structure unadjusted" GWAS analysis ($p=1.27x10^{-6}$, odds ratio 9.55), was only associated with cAD in UK GSDs in the validation study and the level of significance was lower than that seen in the GWAS. This suggests that the result could have been a false positive potentially caused by unrecognised population structure.

The majority of the associations found following validation of SNPs selected from the "population structure adjusted" GWAS analysis were intergenic and the p values were more

modest than found with the "population structure unadjusted" GWAS selected SNPs. However, it has been stated that false positive associations due to population structure can mask actual results[340]. This may be the case when comparing validation of SNPs selected from "population structure unadjusted" and "adjusted" analyses.

In Shiba Inu's a strong association was found with RS23767031. This SNP is upstream of U4 (ENSCAFG00000021950), a small nuclear RNA (snRNA) involved in RNA splicing. In addition, RS24092655, an intronic SNP within Solute carrier family 35 member F3 (SLC35F3), was also associated with cAD. SLC35F3 is thought to be a putative solute transporter[361]. It is therefore possible that it affects Ca^{2+} transport and that it could have implications for keratinocyte differentiation and presence of Filaggrin in keratinocytes[362].

A weak association was found with two intronic SNPs: RS24275059 and RS24354997. These are located within ENSCAFG00000015477, an orthologue of human Radial spoke head 10 homolog B (RSPH10B), the function of which is unknown. This association was found in UK GSDs and UK Labradors (rs24275059 only).

From the GWAS, there were 5 SNPs identified by both "population structure unadjusted" and "adjusted" analyses. Two of these SNPs showed an association in this study: RS22864357, a synonymous coding SNP within Leucine-rich repeats and immunoglobulin-like domains of protein 1 Precursor (LRIG1 [ENSCAFG00000006680]); and RS24332727 within intronic SDK1 (ENSCAFG00000016192) Protein sidekick-1 Precursor. LRIG1 acts as a negative feedback regulator of receptor tyrosine kinase signalling[363]. SDK1 produces a cell adhesion protein which functions in neuronal development[364]. BDNF has been implicated in AD via the itch scratch response[365] and it is possible that a neuronal component is involved in cAD. However, these associations were only seen in UK golden retrievers, and the small numbers within this group make a false positive result a possibility.

The strongest associations in this study were found with SNPs selected from the GWAS "population structure unadjusted" results. This raises questions regarding the suitability of IBS clustering and MDS as a method for defining population structure in dogs. One reason why the "population structure adjusted" analysis may not have identified SNPs which have been validated from the "population structure unadjusted" analysis is that the validated SNPs may be in LD with other SNPs. This would mask their effect in the "population structure adjusted" analysis. This may have happened, as high LD SNPs need to be removed from the analyses to allow population structure to be accurately defined (discussed in Chapter 7). This is difficult in dogs because of the lack of LD information. Furthermore the majority of SNPs in the dog genome appear to be in high LD. This throws into question the use of 'human' population structure adjustment methods for analysis of dog data.

It can also be argued that the meta-analysis clusters used to analyse the Sequenom data may not have been accurate, leading to false positives in the Sequenom validation analysis. The clusters were defined by the user based on the breed and place of origin supplied by the veterinary practice, which could lead to mistakes. Moreover, as seen in Figure 7.3, Japanese and American populations of golden retrievers do not cluster according to origin. This undermines the use of non-marker informed clusters and suggests that false positives due to population structure could be present in the Sequenom meta-analysis results. Ideally the meta-analysis should be repeated with IBS clusters defined from a set of population informative genetic markers run on the entire sample set. This would provide confidence that the population structure effects had been accounted for.

SNPs selected from "population structure unadjusted" and "adjusted" analyses have proven useful in identifying SNPs for further research. The validation of some of the GWAS results does indicate that the use of GWAS as a novel SNP discovery tool is valid in dogs.

For future studies it is recommended that sample selection for GWAS is based on the use of ancestry informative markers to initially define population structure and then select grouped individuals for inclusion on GWAS[334]. Furthermore, using ancestry informative markers on all individuals selected for the large-scale validation study would improve the analysis and confidence in the results.

It may have been beneficial to use the same breed for the validation study to ensure a more direct comparison. The cAD phenotype is clinically heterogeneous[7], which may have caused problems in the across breed comparisons. It is possible that cAD is caused by different loci in different breeds, and in support of this, many of the associations found were only in one breed. However, the fact that the SNP is associated in one breed but not another does not necessarily mean the SNP isn't a risk factor for cAD, as the SNP may have different levels of penetrance in different breeds[366]. There are examples of this in human GWAS where discovery of important SNPs initially made in non-European populations was then demonstrated to be present in European populations at a different penetrance[367-369]. In support of the decision to include multiple breeds for the validation study, it has been stated that although effects of certain variations may be different in each population it can still be beneficial to analyse multiple populations to increase statistical power[366]. Demonstrating the same association in different populations is also advantageous to further study design and increases confidence in the results[326].

The majority of associations found were SNPs in intergenic regions, and, although these were not discussed in detail, they do represent important areas for further fine mapping and re-sequencing. The reason why so many of the associations were within intergenic regions is unknown but it is possible that SNPs could be in linkage disequillibrium with other structural variations, e.g. microsatellites, Variable Number Tandem Repeats (VNTRs), retro-element

insertions, deletions and duplications. It is also important to note that GWAS is limited to identifying common SNPs with small effects; it cannot identify rare SNPs with large effects, un-typed SNPs and parent of origin effects, which could be important in cAD.

In conclusion, this is the first GWAS on cAD that has been subsequently validated in a larger cohort. In total, 30 SNPs were shown to be associated with cAD, 2 of which were associated in all 8 breeds tested. However, methods for analysis of GWAS results in the dog still require some development to ensure population structure is adequately controlled for and to increase the likelihood of identifying lower penetrance alleles.

Chapter 9

Candidate Gene Case-Control SNP Association Study

9. Candidate Gene Case-Control SNP Association Study

9.1 Introduction

In the previous chapter, SNPs selected by GWAS were validated for an association with cAD. In this chapter, SNPs selected from literature searches and following canine microarray (Chapter 3), some of which have been investigated at the expression level by qPCR (Chapter 5), and other SNPs identified by WAVE SNP discovery (Chapter 6) will be assessed for any cAD association using the same panel of dog cases and controls and Sequenom genotyping methodology as in Chapter 8.

This approach is referred to as a candidate gene study (see Chapter 1 for details) and is commonly used to study complex diseases. The basis of this type of study is the same as GWAS, in that if a polymorphism is identified at a higher (or lower) frequency in cases than in controls it may be associated with the disease. The main differences between the approaches are the methods used for SNP selection. GWAS is an unbiased whole genome approach but a candidate gene study is biased as SNPs are selected based on previous evidence of association or relevant function to the disease pathogenesis. Therefore, this type of study is limited to small regions of genes that have already been annotated or previously linked with the disease. It is, nevertheless, relevant to use this approach, as previously discussed the GWAS array used to investigate cAD has gaps in coding regions potentially missing important genes, e.g. Filaggrin, which may be involved in the disease pathology.

Ninety seven SNPs from 25 genes (Table 9.1), all of which have shown previous evidence of functionality or changed expression in cAD have been included in this study. A further 18 SNPs from 10 genes selected from the GWAS data were also included (see Table 9.1). These genes were not covered in the GWAS array but SNPs in intergenic regions

surrounding the genes were (see Figure 9.1). When these intergenic SNPs showed high odds ratios or increased significance in p value, the gene they flanked was included in this study (assuming the gene was not covered by the array). This is similar to the principle of tagging SNPs, although no LD information is available for the canine genome and without this the validity of this method is questionable. See Figure 9.1 for further explanation. Figure 9.2 shows the selection of candidate genes for the genotyping study relative to the rest of this thesis. Table 9.1 lists all genes included in this study and the evidence for inclusion. It was hoped that by including genes with relevant functions to AD or previous linkage/association, an association with cAD would be found.

Figure 9.1 Significant hits in intergenic areas surrounding genes

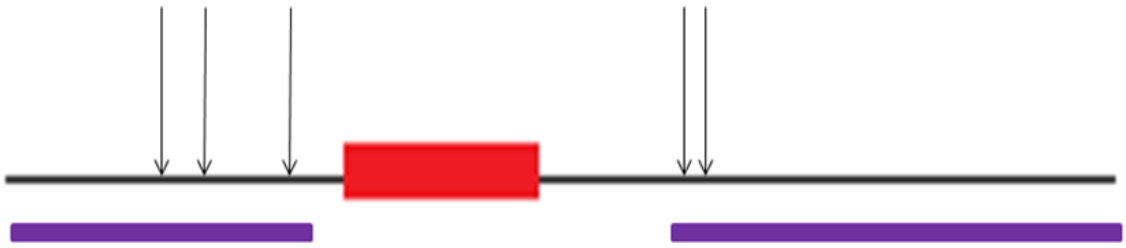


Figure 9.1: The purple area represents the coverage of the Illumina array, the red area is a protein coding gene. It can be seen that the gene is not covered by the array. The arrows represent intergenic SNPs which show high odd ratios or increased significance. In this case it can be seen that these flank the gene, suggesting that the gene could be of importance.

Figure 9.2 Gene selection for Candidate gene case control study

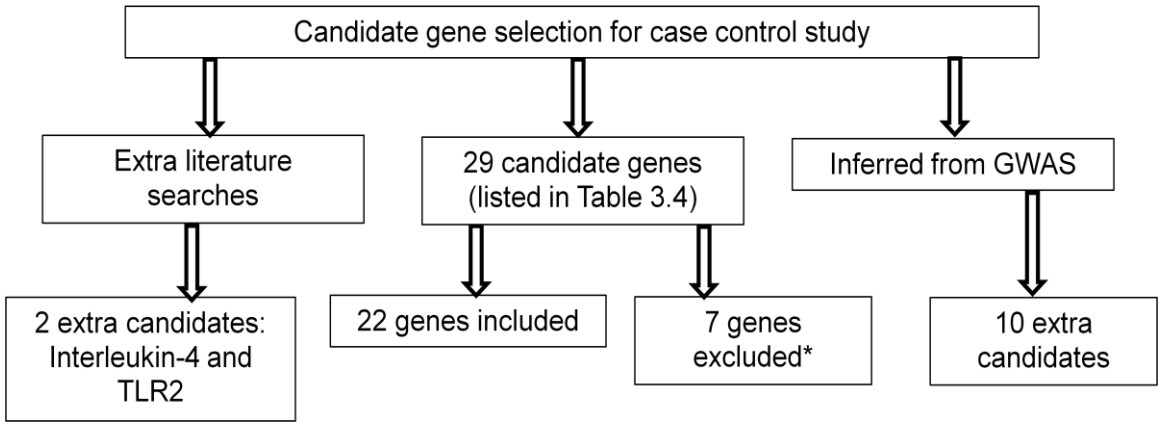


Figure 9.2: Candidate gene selection for this study was based on the original 29 candidate genes listed in Table 3.4, 22 were selected and 7 excluded due to primer design issues. The excluded genes were: GOLGA4, MUC2, IL1RL1, CTLA4, Cystatin A, SOCS3 and phospholipase C, zeta 1. Extra literature searches were undertaken and 2 more candidate genes included; IL-4 and TLR2. 10 genes were inferred from GWAS and included as per Figure 9.1.

Table 9.1 Genes selected for Sequenom study and supporting evidence for inclusion

Gene	Chromosome	Evidence
Interleukin 4	11	Human and canine Literature
ARTS-1	3	Canine microarray
Cadherin 13	5	Canine microarray
POSTN	25	Canine microarray
SYND1	17	Canine microarray
INPPL1	21	Canine microarray and qPCR
SAA-1	21	Canine microarray and qPCR
SCCA2	1	Canine microarray and qPCR
PKP2	27	Canine qPCR
PPAR GAMMA	20	Canine qPCR
P-selectin	7	Canine qPCR and literature
TNF ALPHA	12	Human and canine literature
CARD 4	14	Human Literature
CMA1	8	Human Literature
Filaggrin	17	Human Literature
MS4A2	21	Human Literature
PPAR ALPHA	10	Human Literature
SPRR1B	17	Human literature
TLR2	15	Human literature
TSLP receptor	X	Human literature
STAT2	10	Human literature and canine microarray
DPP4	36	Human Literature and canine qPCR
SPINK5	2	Human Literature and canine qPCR
S100A8	7	Human literature, canine microarray and canine qPCR
CUX1	6	Based on region identified by GWAS
DOCK2	4	Based on region identified by GWAS
GATA-3	2	Based on region identified by GWAS
GTF2I	6	Based on region identified by GWAS
IRAK1BP1	12	Based on region identified by GWAS
NEK7	7	Based on region identified by GWAS
PDGFD	5	Based on region identified by GWAS
SORCS2	3	Based on region identified by GWAS
TFEB	12	Based on region identified by GWAS
COL19A1	12	Based on region identified by GWAS

9.2 Methods

Atopic dogs and healthy controls – breed, sex and age matching

The same samples were used as in Chapter 8 (see section 8.2)

Primer and probe design

See Appendix 12 for full primer and probe sequences for the SNPs listed in Table 9.2.

Table 9.2 lists the SNPs included in the Sequenom genotyping assay.

Table 9.2 SNPs included in the Sequenom genotyping assay.

RS number	Chr	Base Pair	Gene	Location
rs21902676	1	16469324	SCCA2	Upstream
rs21941198	1	16475961	SCCA2	Non synonymous coding
rs21902632	1	16476102	SCCA2	Non synonymous coding
rs9013510	1	16476441	SCCA2	Downstream
rs21902631	1	16476541	SCCA2	Downstream
rs9013509	1	16476584	SCCA2	Downstream
rs22821176	2	30911977	GATA-3	Downstream
rs22750909	2	30930868	GATA-3	Upstream
rs22871624	2	45133450	SPINK5	Upstream
rs22871627	2	45133465	SPINK5	Upstream
rs22871630	2	45133611	SPINK5	Upstream
rs22871646	2	45133786	SPINK5	Upstream
rs22819627	2	45159024	SPINK5	Non synonymous coding
rs22872666	2	45162084	SPINK5	Synonymous coding
rs22818977	2	45167469	SPINK5	Synonymous coding
rs22874976	2	45176838	SPINK5	Splice site, intronic
rs22839343	2	45201929	SPINK5	Synonymous coding
rs22807086	2	45212191	SPINK5	Downstream
rs23593765	3	15767342	ARTS-1	Synonymous coding
rs9196801	3	15767484	ARTS-1	Frame shift - coding
rs24872954	3	15782443	ARTS-1	Synonymous coding

Table 9.2 continued

rs9190946	3	15784064	ARTS-1	Synonymous coding
rs8857124	3	62477640	SORCS2	Synonymous coding
rs9241373	3	62483849	SORCS2	Frame-shift coding, splice site
rs23586992	3	62483869	SORCS2	Synonymous coding
rs24150067	4	45333435	DOCK2	Synonymous coding
rs24202381	5	31088631	PDGFD	Synonymous coding
rs24264853	5	71503539	Cadherin-13	Downstream
rs8988459	6	8748612	GTF2I	Synonymous coding
rs9255897	6	10862928	CUX1	Synonymous coding
rs24328835	6	10890690	CUX1	Synonymous coding
rs8667691	6	10898270	CUX1	Synonymous coding
rs24319791	6	10917564	CUX1	Synonymous coding
rs24404799	7	7553657	NEK7	Non synonymous coding
rs24404800	7	7553658	NEK7	Non synonymous coding
rs24380950	7	7570280	NEK7	Synonymous coding
rs24435410	7	31958066	P-Selectin	Non synonymous coding
rs24438947	7	46517606	S100A8	Upstream
s100a8_p	7	46517638	S100A8	Promoter region
s100a8_p2	7	46517792	S100A8	Promoter region
s100a8_p3	7	46517859	S100A8	Promoter region
rs8867188	8	7447456	CMA1	Upstream
rs22655272	9	5852692	SOCS3	Upstream
rs22655274	9	5852893	SOCS3	Upstream
rs22655276	9	5853219	SOCS3	Upstream
rs22655278	9	5853780	SOCS3	Upstream
rs22615134	9	5860501	SOCS3	Downstream
rs22000453	10	3625056	STAT2	Downstream
rs22000474	10	3625180	STAT2	Downstream
rs22029672	10	3631328	STAT2	Non synonymous coding, splice site
rs22009467	10	23005117	PPAR-α	Splice site, intronic
rs22146864	11	23976639	IL-4	Upstream
rs9095590	12	4078693	TNF-α	Synonymous coding

Table 9.2 continued

rs22217971	12	13373163	TFEB	Synonymous coding
rs22263793	12	35504037	COL19A1	Synonymous coding
rs8860216	12	42878364	IRAK1BP1	Downstream
rs22374481	14	46159795	CARD4	Upstream
rs22374487	14	46159824	CARD4	Upstream
rs22410121	15	54484550	TLR2	Non synonymous coding
rs22349093	15	54485784	TLR2	Synonymous coding
rs22349092	15	54486859	TLR2	Downstream
rs22605727	17	18202904	SYND1	Downstream
rs22529256	17	64273415	Filaggrin	Synonymous coding
rs22542234	17	64296141	Filaggrin	Non synonymous coding
rs8555036	17	64296436	Filaggrin	Synonymous coding
rs22542260	17	64296760	Filaggrin	Synonymous coding
rs22542262	17	64296821	Filaggrin	Non synonymous coding
rs22542263	17	64296837	Filaggrin	Splice site, intronic
rs22588226	17	64297000	Filaggrin	Non synonymous coding
rs22588227	17	64297022	Filaggrin	Non synonymous coding
rs22588228	17	64297028	Filaggrin	Non synonymous coding
rs8555045	17	64298442	Filaggrin	Non synonymous coding
rs8555046	17	64298495	Filaggrin	Non synonymous coding
rs8555044	17	64298566	Filaggrin	Synonymous coding
rs22511502	17	64298621	Filaggrin	Non synonymous coding
rs22511526	17	64298647	Filaggrin	Non synonymous coding
rs22509070	17	64298685	Filaggrin	Non synonymous coding
rs22511501	17	64300264	Filaggrin	Synonymous coding
rs22509072	17	64300322	Filaggrin	Non synonymous coding
rs22509074	17	64300703	Filaggrin	Non synonymous coding
rs22511518	17	64301581	Filaggrin	Synonymous coding
rs22542266	17	64301700	Filaggrin	Non synonymous coding
rs22511504	17	64301968	Filaggrin	Synonymous coding
rs22509090	17	64301986	Filaggrin	Synonymous coding
rs22509067	17	64302067	Filaggrin	Non synonymous coding

Table 9.2 continued

rs22565034	17	64303134	Filaggrin	Non synonymous coding
rs22599138	17	64303187	Filaggrin	Non synonymous coding
SPRR1B_e3_1	17	64883377	SPRR1B	Exon 3
rs8944179	20	9156340	PPARγ	Synonymous coding
rs22931896	20	9213048	PPARγ	Upstream
rs22931897	20	9213053	PPARγ	Upstream
rs22935368	21	28904622	INPPL1	Downstream
rs9070190	21	28910870	INPPL1	Synonymous coding
rs9150577	21	28912392	INPPL1	Synonymous coding
rs22976618	21	28912506	INPPL1	Synonymous coding
rs8488679	21	43871640	SAA-1	Upstream
rs22908459	21	43872759	SAA-1	Synonymous coding
rs22910701	21	43875650	SAA-1	Synonymous coding
rs22908463	21	43891974	SAA-1	Synonymous coding
rs22986026	21	53629181	MS4A2	Upstream
rs8877519	21	53631150	MS4A2	Splice site, synonymous coding
rs8877521	21	53631234	MS4A2	Synonymous coding
rs22986033	21	53631825	MS4A2	Non synonymous coding
MS4A2_E5	21	53634435	MS4A2	Exon 5
rs22986082	21	53634938	MS4A2	Synonymous coding
rs9019919	21	53636716	MS4A2	Downstream
rs23274229	25	6184255	POSTN	Synonymous coding
rs23360288	27	19048662	PKP2	Synonymous coding
rs23907319	36	10444284	DPP4	Synonymous coding
rs24630571	X	745241	TSLP receptor	Synonymous coding
rs24630609	X	745403	TSLP receptor	Synonymous coding
rs24613434	X	750979	TSLP receptor	Synonymous coding
rs24613432	X	751050	TSLP receptor	Splice site, intronic
rs24639522	X	755348	TSLP receptor	Synonymous coding
rs24657063	X	755447	TSLP receptor	Synonymous coding

Sequenom genotyping

97 SNPs had probe and primer sequences successfully designed for Sequenom-based genotyping. These SNPs were genotyped in each dog using the Sequenom platform.

See Chapter 2 for details of quality control, plexing and plate setup for the Sequenom. The methods used for Sequenom genotyping were the same as in Chapter 8. See Chapter 8 methods (section 8.2) for details of PCR, SAP treatment, iPLEX reaction, resin clean, spotting and running of the arrays. The data collection, calling and analysis were also performed as detailed in Chapter 8 (section 8.2).

Statistical analysis

The whole-genome association analysis toolset PLINK was used to analyse the data. All data were subject to quality control procedures; only SNPs which conformed to HWE, exhibited ≥90% genotyping success and had a minor allele frequency≥0.01 were included in subsequent analysis.

As in Chapter 8, Figure 8.1 three types of analysis were conducted. Asssociation tests were undertaken on all breeds separately, including basic association, allelic association, logistic regression and the genotypic model statistical test. Meta-analysis was also undertaken on all breeds and individual breeds grouped according to geographic region. This considers the data as a whole taking into account variables between them such as breed and geographic region (defined by user from supplied phenotype information), eliminating false positives due to population substructure. P values were corrected for multiple hypotheses testing by multiplying p values by the number of independent loci (48 in this study).

9.3 Results

Table 9.3 shows the SNPs found to be cAD associated in all eight dog breeds tested using meta-analysis. In each sample, the corrected p value represents the level of significance after correction for multiple testing, which is based on the number of independent loci in the analysis (48 in this study). Only one SNP was significant after correction, this was rs24613432 which is intronic within TSLP receptor.

Table 9.4 shows the results from the individual breed analysis as a matrix of corrected p values obtained from the allelic association tests. Where indicated, the cAD associations were significant on the logistic and genotypic association tests (See Appendix 13 for a more complete summary of the results). The results from the individual breed meta-analysis are not shown in the main text as there were no significant results after multiple testing correction (see Appendix 14 for a summary of the results). 15 SNPs were associated with cAD, although these associations were in a variety of dog breeds.

Table 9.3 Meta-analysis of All Breeds Clustered by Breed and Geographic region

SNP	CHR	BP	Gene	Location	Odds ratio	P value	Corrected p value
RS24613432	X	751050	TSLP receptor	Splice Site, Intronic	0.38	8.0×10^{-4}	0.037
RS23907319	36	10444284	DPP4	Synonymous Coding	1.54	0.004	0.20
SPRR1B_e3_1	17	64883377	SPRR1B	Exon 3	0.64	0.011	0.51
RS22819627	2	45159024	SPINK5	Non Synonymous Coding-Amino Acid Change D/E	1.40	0.014	0.67
RS21902632	1	16476102	SCCA2	Non Synonymous Coding – Amino Acid Change R/S	0.70	0.017	0.84
RS8860216	12	42878364	IRAK1BP1	Downstream	1.66	0.023	1.08
RS22565034	17	64303134	Filaggrin	Non synonymous Coding – Amino Acid Change E/K	0.36	0.026	1.23
RS22818977	2	45167469	SPINK5	Synonymous coding	0.75	0.031	1.48
RS9150577	21	28912392	INPPL1	Synonymous coding	0.76	0.033	1.56
RS9190946	3	15784064	ARTS-1	Synonymous coding	1.33	0.043	2.08
RS8877521	21	53631234	MS4A2	Synonymous coding	1.28	0.044	2.10

Table 9.4 Individual Breed Analysis: Matrix of Corrected P value Significant SNPs in cAD

		Boxer		**GSD**			**Labrador**			**Golden retriever**		**Shiba Inu**	**Shih Tzu**	**Pit Bull**	**WHWT**		
Rs number	*Gene region*	*UK*	*All*	*UK*	*USA*	*All*	*USA*	*UK*	*All*	*USA*	*UK*	*Japan*	*Japan*	*USA*	*All*	*UK*	*USA*
RS22009467	PPAR-alpha	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.031
RS22542260	Filaggrin	-	-	-	-	-	-	-	-	-	0.007*	-	-	-	-	-	-
RS22542263	Filaggrin	-	-	-	-	-	-	-	-	-	0.014*	-	-	-	-	-	-
RS22588227	Filaggrin	-	-	-	-	-	-	0.009**		-	-	-	-	-	-	-	-
RS22807086	SPINK5	-	-	-	0.014	-	-	-	-	-	-	-	-	-	-	-	-
RS22871630	SPINK5	-	-	-	-	-	-	-	-	0.021[a]	0.030*	-	-	-	-	-	-
RS22986026	MS4A2	-	-	-	-	-	-	-	-	-	-	8.7 x 10-4*	-	-	-	-	-
RS23274229	POSTN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.001
RS23907319	DPP4	-	-	-	-	-	-	5.2 x10-4**		-	-	-	-	-	-	-	-
RS24319791	CUX1	-	-	-	-	-	-	-	-	-	-	-	-	0.044	-	-	-
RS24328835	CUX1	-	-	-	-	-	-	-	-	-	-	0.012	-	-	-	-	-
RS24613432	TSLP receptor	-	-	-	-	-	-	-	-	-	-	-	-	0.005*	-	-	-
RS24613434	TSLP receptor	-	-	-	-	-	-	-	-	-	0.020	-	-	-	-	-	-
RS9070190	INPPL1	-	-	-	-	-	-	-	-	-	-	4.6 x10-5**	-	-	-	-	-
RS9150577	INPPL1	-	-	-	-	-	-	-	-	-	-	1.9 x10-4**	-	-	-	-	-

* denotes that this association is significant on trend analysis using the genotypic model. ** denotes that this association is significant on trend analysis using the genotypic model and logistic association using the additive model. [a] denotes that this association did not make the HWE cut off of 0.05

9.4 Discussion

As these data include multiple breeds, the analysis needs to be carefully selected. Meta-analysis allowed all of the data to be clustered according to breed and point of origin information and then analysed independently of these variables. Hence, meta-analysis was the preferred method for analysis of this type of data, giving greater statistical confidence due to the larger number of cases and controls. It was, however, useful to analyse these data breed by breed as there was no assurance that the same disease phenotype in different breeds had the same genetic cause. Moreover, in complex diseases, such as AD, different levels of penetrance can be observed in different SNPs and breeds.

Data derived from this study clearly showed some SNP associations with cAD. It was also clear, from a visual check of the data sets, that these SNP associations were different from breed to breed. In fact, only one SNP was associated with cAD in all 8 breeds. However up to 4 SNPs were found to be cAD associated in some breeds when studied individually.

The results from the meta-analysis on all eight breeds revealed only one SNP (rs24613432) to be significantly associated with cAD after correction for multiple testing. This SNP is within Thymic stromal lymphopoietin (TSLP) receptor gene. The same SNP was also significantly associated in the Pit Bull terrier when using allelic association testing and the genotypic model (Table 9.4). RS24613434 is another SNP within TSLP receptor that was found to be significant in UK and European golden retrievers (Table 9.4).

TSLP receptor forms a functional complex with TSLP and IL-7 receptor which can stimulate cell proliferation through STAT3 and STAT5 activation[370]. It can also activate JAK2 and is expressed by dendritic cells and activated monocytes[371]. Its involvement in the JAK-STAT pathway[217], which is used by interleukins to transduce the signals that initiate cytokine specific gene transcription, is of interest in the immunological response of AD.

Furthermore, the functional complex with TSLP links its function closely with that of TSLP, which has been called the 'master switch' for allergic inflammation[182]. Evidence for this is based on observations of TSLP-treated human blood dentritic cells which induce an inflammatory Th2 cytokine pattern[239]. Increased expression of TSLP by keratinocytes in atopic skin lesions in hAD has been observed by immunohistochemistry[239]. It has also been demonstrated that TSLP "converts human epidermal Langerhans cells into antigen-presenting cells that induce pro-allergic T cells"[242]. This is of interest because previous studies have noted that Langerhans cells serve as antigen-presenting cells in AD[243;244], by expressing high-affinity IgE receptors that allows for more efficient allergen capture, processing and presentation to T cells[245]. This effect of TSLP could render LCs capable of inducing or maintaining atopic dermatitis symptoms[242].

TSLP, however, is yet to be characterised in the canine genome, consequently it could not be included here, although the TSLP receptor was included. Because of TSLP's potential relevant function to the pathogenesis of AD it is likely that TSLP receptor may also have a role. As these results demonstrate a cAD association with TSLP receptor, they indicate that it is a good candidate for further fine mapping and re-sequencing studies.

The remaining single breed analysis presented in Table 9.4 found 15 SNPs in nine genes to be significantly associated with cAD in a variety of breeds. CUX1 was one of these genes, with associations being found in the Pit Bull terrier and the Shiba Inu. These associations were found in different SNPs, rs24319791 and rs24328835 respectively. Both of these are synonymous coding suggesting that a non-synonymous SNP that was not included in the Sequenom assay could be the actual causative SNP. It is interesting to note that CUX1 was not selected for Sequenom analysis based on its function or previous association with hAD, but was inferred from the GWAS conducted in Chapter 7. Suggesting that inference from

GWAS as described in the introduction may be a useful way to compensate for the gaps present on the canine Illumina SNP20 chip (discussed in Chapter 7). Of further note is the interaction of CUX1 with GOLGA5[372], a mRNA transcript that was found to be dysregulated on the canine microarray performed in Chapter 3.

Two SNPs in SPINK5 were found to be associated with cAD: rs22807086 in USA GSDs, and rs22871630 in UK and USA golden retrievers although the association with USA golden retrievers did not met the HWE cut-off. In the meta-analysis two SPINK5 SNPs (rs22819627 and rs22818977) showed an association with all eight breeds, but this was not significant after multiple testing corrections. This is noteworthy because these multiple weak associations with SPINK5 suggest a link with this gene and the disease. It is possible that the causative SNP was not included in this study or different SNPs within this gene lead to the same phenotype in different breeds. Further analysis by fine mapping and re-sequencing will be required to better characterise the SNPs in this gene and locate the causative mutation. SPINK5 is a good candidate for further research because of these associations and also because it functions in hair and skin morphology, proteolysis during epithelial formation and keratinocyte differentiation[170]. Moreover, as discussed in Chapter 3, SPINK5 mutations have been shown to be associated with hAD[125-127] and in Chapter 5 differential gene expression of SPINK5 in cAD was demonstrated.

Peroxisome proliferator-activated receptor alpha (PPARα, rs22009467) was associated with cAD in the WHWT. Members of the PPAR family (PPARβ/δ, PPARα, PPARγ) have been associated with inflammation and skin barrier function in humans, and have been implicated in the pathogenesis of hAD[209]. PPARs represent interesting candidates because treatment with them has improved atopic symptoms in human patients[209]. However, the numbers of WHWT tested were low, meaning that this may be a false positive result.

In hAD, two Filaggrin loss of function variants (R501X and 2282del4) were identified to be major risk factors for the development of AD[51;52;228]. Filaggrin contributes to barrier formation by causing the cells to compact into squames by aggregating to the keratin cytoskeleton. It can also be broken down at the stratum corneum to form a "natural moisturizing substance"[42;53]. In this study, the UK golden retrievers showed an association with Filaggrin SNPs (rs22542260 and rs22542263), although the numbers of dogs were low, suggesting that these results may be false positives. Nevertheless, the UK Labradors showed an association with rs22588227, which is a non-synonymous SNP in Filaggrin, and therefore, may lead to a loss of function. In addition, rs22565034, another non-synonymous SNP, showed an association on the meta-analysis, although this was not significant after correcting for multiple testing. The weak association on meta-analysis could be due to differing SNPs within the gene leading to the same phenotype in different breeds. Also, the associations found here could represent SNPs in linkage disequillibrium with another causative SNP not included on the array. There were many non-synonymous SNPs in Filaggrin that were not included on this array, due to assay design issues, that could be contributing to the disease phenotype.

Dipeptidyl-peptidase 4 (DPP4) (rs23907319) was found to have a strong association (confirmed by logistic association and genotypic model tests) with cAD in UK Labradors. The same SNP was identified on meta-analysis as potentially cAD associated, although it was not significant after correction for multiple testing. DPP4 has been implicated in: activation of regulatory T cells, T cell cycle arrest, inhibition synthesis and proliferation of T cells, suppression of inflammatory cytokines by T cells[191;192] and wound healing[193]. In Chapter 5, mRNA expression of DPP4 was down-regulated in cAD and this, coupled with the association evidence here, suggests that it is a good candidate for further re-sequencing and fine mapping research.

MS4A2 produces FcεRIβ, which is important in IgE function[43;112]. As discussed in Chapter 1, FcεRIβ binds allergen specific IgE expressed on Langerhans and mast cells[32], leading to the inflammatory responses seen in allergic inflammation. One MS4A2 SNP, rs22986026, was associated with cAD in the Shiba Inu (rs8877521 was also associated on the meta-analysis, but after correction for multiple testing it was not significant). The functions described and associations seen in this experiment do indicate that this gene is of interest in understanding the pathogenesis of cAD.

Two SNPs in INPPL1 (rs9070190 and rs9150577) were strongly associated (confirmed by logistic association and genotypic modelling) with cAD in the Shiba Inu. Another SNP, rs9150577, was associated in meta-analysis but was not significant after correction. INPPL1 was up-regulated in cAD in the qPCR experiment described in Chapter 5. The encoded protein also plays a role in the regulation of epidermal growth factor and controls the shift of lipid balance by increased tyrosine phosphorylation[373]. Moreover, the SRC2 domain present on INPPL1 has been implicated in T cell development[278]. This evidence of T cell and lipid involvement, along with potential disease associations, strengthens the evidence for a pathogenetic role for this gene in cAD.

An association between USA WHWTs and POSTN (rs23274229) was seen, although the case numbers were small, therefore this could represent a false positive result.

Finally, SNPs in SPRR1B, SCCA2 and IRAK1BP1 were found to be associated on meta-analysis but not after correction. The fact that these genes did not stand up to multiple testing and were not implicated on any of the other association tests suggests they may be false positives.

There are issues to consider with this study. Firstly as there were eight different breeds from variable locations, the numbers for single breed analysis were small, especially for a

complex disease such as cAD. It is possible that associations may have been missed due to a lack of statistical power. Repeating this experiment with larger numbers and better defined breeds from one area (e.g. the UK) could strengthen the results. Moreover, cAD is a clinically heterogeneous condition that may have several different genetic determinants impacting upon the clinical outcome. Different breeds and different individuals within a breed may also have different genetic determinants leading to the same clinical phenotype. This is possible because SNPs in linkage disequilibrium with each other can be variable in different breeds. For example, multiple associations were seen within Filaggrin and SPINK5 in different breeds indicating that a shared causative SNP may have not been genotyped. Further fine mapping and re-sequencing of these areas may identify a causative mutation common to all breeds that was not included in this study. In addition, multiple associations within the same gene may indicate a haplotypic effect. Haplotype analysis, however, could not be performed on these results because the data was not genome wide or of sufficient density. If the genes indicated here were to be fine mapped this would allow for haplotype analysis of that area. Furthermore, LD analysis could not be undertaken because dogs do not have a HAPMAP equivalent. Nevertheless, this study does illustrate some potential associations with good candidate genes which warrant further research by fine mapping and re-sequencing.

Chapter 10

General Discussion

10. General Discussion

This thesis set out to investigate the genetics of cAD and to assess whether cAD was a suitable model for the human disease. In order to achieve this, a candidate gene approach was used. Selection of candidate genes was based on a canine mRNA microarray, qPCR and literature searches. SNPs were selected from these candidate genes to be included in a large-scale case-control comparison. Further to this a GWAS was undertaken which identified potential associations with novel SNPs, and these associations were validated in a large-scale case-control comparison study. Using a variety of approaches to select SNPs for study helped to reduce bias and include novel candidates. In order to test the associations of cAD with these SNPs the genotyping platform Sequenom was used. In total 659 dogs across 8 breeds were genotyped in 232 SNPs across 54 genes and 41 intergenic regions. From this 45 putative associations were found in various breeds (summarised in Table 10.1).

Table 10.1 Summary of the Associations Found by Sequenom

RS Number	Gene/Region	Breed associated	Study
RS9150577	INPPL1	Shiba Inu	Candidate Gene Study- Canine microarray (chapter 3)
RS9070190	INPPL1	Shiba Inu	Candidate Gene Study- Canine microarray (chapter 3)
RS23274229	POSTN	USA WHWT	Candidate Gene Study- Canine microarray (chapter 3)
RS24319791	CUX1	Pit Bull	Candidate Gene Study- Inferred from GWAS
RS24328835	CUX1	Shiba Inu	Candidate Gene Study- Inferred from GWAS
RS24613432	TSLP receptor	All Breeds	Candidate Gene Study- Literature
RS24613434	TSLP receptor	UK golden retriever	Candidate Gene Study- Literature
RS22542260	Filaggrin	UK Golden retriever	Candidate Gene Study- Literature
RS22542263	Filaggrin	UK Golden retriever	Candidate Gene Study- Literature
RS22588227	Filaggrin	UK Labrador	Candidate Gene Study- Literature
RS22986026	MS4A2	Shiba Inu	Candidate Gene Study- Literature
RS22009467	PPAR-alpha	USA WHWT	Candidate Gene Study- Literature
RS23907319	DPP4	UK Labrador	Candidate Gene Study- Literature and canine qPCR (chapter 5)
RS22807086	SPINK5	USA GSD	Candidate Gene Study- Literature, canine qPCR (chapter 5)
RS22871630	SPINK5	USA & UK golden retriever	Candidate Gene Study- Literature, canine qPCR (chapter 5)
RS22864357	LRIG1	UK golden retriever	GWAS Validation Study
RS23602938	PROM1	All golden retrievers	GWAS Validation Study
RS22859255	Rab-3C	All golden retriever	GWAS Validation Study
RS22915894	Rab-7a	UK Labradors & UK golden retriever	GWAS Validation Study
RS24332727	SDK1	UK golden retriever	GWAS Validation Study
RS24092655	SLC35F3	Shiba Inu	GWAS Validation Study
RS24872415	SORCS2	UK GSD	GWAS Validation Study
RS24275059	ENSCAFG00000015477	UK GSD & UK Labrador	GWAS Validation Study
RS24354997	ENSCAFG00000015477	UK GSD	GWAS Validation Study
RS23224877	Upstream - ENSCAFG00000007979	UK WHWT	GWAS Validation Study

Table 10.1 continued

RS23767031	Upstream - U4	Shiba Inu	GWAS Validation Study
RS22523565	Downstream - KIAA1549	USA golden retriever	GWAS Validation Study
RS8806978	Downstream - RAB7a	UK Labrador, All WHWT & UK golden retriever	GWAS Validation Study
RS22114085	Intergenic - Chromosome 10	All Breeds	GWAS Validation Study
RS22184220	Intergenic - Chromosome 12	USA GSD	GWAS Validation Study
RS22720467	Intergenic - Chromosome 19	UK GSD & UK Labrador	GWAS Validation Study
RS23026134	Intergenic - Chromosome 22	USA Labradors & UK golden retriever	GWAS Validation Study
RS23316823	Intergenic - Chromosome 26	USA golden retriever	GWAS Validation Study
RS23472497	Intergenic - Chromosome 29	All Breeds	GWAS Validation Study
RS9132002	Intergenic - Chromosome 29	UK Labrador	GWAS Validation Study
RS23545080	Intergenic - Chromosome 3	All golden retrievers	GWAS Validation Study
RS23690306	Intergenic - Chromosome 31	UK golden retriever	GWAS Validation Study
RS23708599	Intergenic - Chromosome 31	USA golden retriever	GWAS Validation Study
RS23769251	Intergenic - Chromosome 32	UK GSD	GWAS Validation Study
RS23770102	Intergenic - Chromosome 32	USA golden retriever	GWAS Validation Study
RS23968934	Intergenic - Chromosome 36	UK golden retriever	GWAS Validation Study
RS24194054	Intergenic - Chromosome 5	UK WHWT	GWAS Validation Study
RS24327271	Intergenic - Chromosome 6	All golden retrievers	GWAS Validation Study
RS24408651	Intergenic - Chromosome 7	All golden retrievers	GWAS Validation Study
RS24482628	Intergenic - Chromosome 8	All WHWT	GWAS Validation Study

Interesting candidate genes were found to be associated with cAD by Sequenom genotyping. SPINK5 and TSLP receptor in particular were associated with cAD in certain breeds indicating similarities to the human condition, and opening cAD research to further functional study. The GWAS that was undertaken was the first study of cAD using a genome wide SNP array. Novel areas were identified by GWAS and confirmed by Sequenom – two SNPs in intergenic regions showed independent and strong associations with all breeds. However an association with intergenic regions does raise some questions as to how these could be affecting the disease phenotype. It is possible that other variations, not SNPs, could be effecting the association, for example copy number variations, imprinting/methylation, long range genetic regulation via promoter regions and other structural variations, e.g. microsatellites, Variable Number Tandem Repeats (VNTRs), retro-element insertions, deletions and duplications. This is possible because the SNP could be in linkage disequilibrium with the variation leading to increased association with that SNP on GWAS despite it not being the disease causative allele. There are many different types of variation that can contribute to disease pathogenesis, for example increased copy number of β defensin is associated with psoriasis[374]. Only one other type of mutation, a microsatellite in one gene, was investigated in this study, as it was outside the scope of this thesis to further investigate other sources of mutation/variation.

These results will lead to further work in fine mapping and re-sequencing of these areas to identify the causative SNPs or the SNPs in linkage disequilibrium with this region.

This study has been reasonably successful in identifying candidate genes and SNP associations in cAD, however there are a number of criticisms that can be made regarding gene selection and assay design. For example in the candidate gene section for this study a lot of the candidate genes were chosen because of dysregulation on canine mRNA microarray

(which was later confirmed using qPCR), but the coverage of the microarray was poor, especially in regions covering Filaggrin. It may have been beneficial to re-analyse the microarray data using Gene Set Analysis (GSA), which treats expression data as a group and measures subtle co-ordinated changes in the gene group[375]. This means that all changes in expression are relevant and no cut-offs are applied to the data (cut-off is usually less than 2 fold change), shifting the emphasis to the small cumulative and potentially co-ordinated changes that may be more representative of what is seen in complex diseases. Network theory[376] supports this approach stating that biological systems function as a fine network with certain genes functioning as hubs which connect to multiple genes[377]. Therefore a small change in one part of the network may lead to subtle changes across the whole network. GSA tools, however, are limited to humans and the major model species because there is more genomic and gene interaction information available. If a GSA tool is created for dogs it is possible that new information could be gained from the microarray data.

If an mRNA expression study was undertaken now, recent developments in technology would have to be taken into account. RNA-seq is a relatively new process that allows for quantitative measurement and accurate sequencing of RNA essentially combining microarray with qPCR[158]. Therefore it would be a more time efficient, unbiased and accurate method than microarray and qPCR confirmation. Furthermore, by combining the analysis of the microarray data and the GWAS data using GSEA-SNP analysis, SNP information can be combined with expression data leading to a pathway driven analysis[378]. This approach is more geared to understanding the biological process as a whole and identifying the pathways interacting in disease pathology. It was not possible to do this analysis as the statistical power would not have been sufficient due to the limited amount of samples used in GWAS and lack of coverage on the canine microarray. Nevertheless this is an interesting approach to analysis of this data type and could prove useful in the future.

The GWAS array was not optimal because of the assay design; there were a limited amount of SNPs and large gaps on the array, meaning that areas such as Filaggrin were not included. Affymetrix has recently produced a larger canine SNP array with better coverage, which may be beneficial in future studies. The statistical analysis of GWAS studies is still developing, even in humans. For example the WTCCC study has been re-analysed in multiple ways giving new information each time[379-382]. It is also important to note that GWAS is limited to identifying common SNPs with small effects, and it cannot identify rare SNPs with large effects, un-typed SNPs, parent of origin effects and, as discussed above, other variations (with the execption of CNVs).

The GWAS data provides Copy Number Variants (CNV) information; CNV's are defined as stretches of DNA larger than 1 kb that display differences in copy number in the population[383]. As they are variable in the population their association with disease has been investigated. Effects of the copy number of individual CNVs have been associated with diseases through dosage effects of a single gene, a neighbouring set of genes as in Williams-Beuren syndrome and infantile spasms or through allele combinations in the case of some complex diseases[384]. CNV analysis can be done on the GWAS data presented in this thesis, but dogs do not have a whole genome CNV catalogue making this process difficult[385]. This may be something to consider for the future because CNVs are increasingly being studied and linked with complex diseases[384].

HAPMAP provides human researchers with a rich data set of linkage and SNP information which is invaluable in interpreting results or calculating frequencies of tagging SNPs. The major limitation with analysis of GWAS data in dogs is the fact that no HAPMAP equivalent exists. A recent study (Ke *etal* "Assessment of the functionality of genome-wide canine SNP arrays and implications for canine

disease association studies," manuscript in prep) noted that because of the lack of a HAPMAP the canine GWAS arrays were designed based on the Boxer assembly and SNPs were selected by physical distance. This means that LD coverage and SNP tagging capability varies greatly from breed to breed. This suggests that the selection of golden retrievers for the GWAS may have been ill advised due to reduced LD and SNP tagging. Also only 48 samples were included due to budget constraints. It would be better in future studies to use a larger number when studying complex diseases to allow low penetrance SNPs to be identified.

Other criticisms of this study include the number of SNPs typed and the case-control selection. SNPs/genes shown to be of great importance in hAD, for example Filaggrin[52;228;230;231;359;360], were not included on the microarray, qPCR or covered in the GWAS. A small number of Filaggrin SNPs were included on the Sequenom, but the majority of SNPs available for Filaggrin could not be included, due to the highly repetitive sequence of Filaggrin which limits assay design[386] for Sequenom, microarray and qPCR. The Filaggrin SNPs that were included on the Sequenom were mainly synonymous SNPs, again because of assay design limitations, rather than the more crucial non synonymous SNPs. However weak associations were found in the Filaggrin SNPs included, and therefore further research by fine mapping and re-sequencing is recommended.

Case and control selection for the genotyping was breed, sex, age and country of origin matched for the most part. But grouping by country of origin may have been flawed as the population structure results in Chapter 7 (Wood *etal* "Genome Wide SNP Arrays Reveal Underlying Population Substructure in domestic Dog Breeds", manuscript in prep) indicates that there are discreet levels of population structure even within a breed from the same point of origin. In human Malaria research similar issues of population stratification effect the result outcome - the populations used are stratified despite the fact they had all been sampled

from similar regions. The stratification was based on the dialect of each village meaning each village and dialect was representing a separate population that was affecting the results[387-389]. This demonstrates that population structure can have a significant effect on the study outcome and highlights the importance of defining population structure accurately. Therefore it would have been better to use ancestry informative markers to group and select dogs for genotyping. This would have made the results more accurate and improve the statistical power. However, it could have also reduced the number of useable samples to low levels and would have exceeded time and financial restraints.

The major limiting factor in this research was the availability of case and control samples. Samples could only be collected if they were excess to diagnostic requirements, therefore the control group often had a disease state. This could have led to segregation in the control group due to other disease states. Efforts to avoid this were made, and samples were generally taken from 'healthy' dogs with no history of immunological or skin disease (although the samples taken from the VLA had no phenotype information available). Collection of cAD samples was also difficult because confirming the diagnosis takes time and the heterogeneity of the clinical presentation may have affected the results.

Future research in cAD should focus around fine mapping and re-sequencing of the areas identified by GWAS and Sequenom. Using a small sample set of severe cases of cAD and fine mapping them may allow rare variants not identified by GWAS to be found. Moreover, there is an opportunity for functional work on some of the regions.

Another potential area for investigation is epigenetics; methylation studies could be undertaken in SPINK5, for example, perhaps identifying maternal inheritance as suggested by human research[170]. A mouse epigenetic study indicated that distal regulatory elements can direct transcription of INF-γ[390]. As INF-γ is important in the pathogenesis of AD

investigating this epigenetic regulation would of interest in cAD. Studies using RNA interference (RNAi) in keratinocyte cultures could also allow the role of particular genes to be identified, although this requires a robust and validated *in vitro* model for cAD. Protein profiling would also be of use in cAD research, either through immunohistochemistry or proteomic approaches.

The findings presented in this thesis are broadly similar to the human condition, indicating that cAD is a good model for hAD. Further work is required to confirm, nevertheless dogs could provide an excellent tool for revealing gene-environment interactions in AD because of their shared environment with humans.

This work has the potential to inform further studies into the pathogenesis of cAD. By understanding the pathology of the disease it is possible that better treatments for cAD can be developed. Furthemore by illustrating similarities between cAD and hAD treatments used for humans could be considered for dogs, for example improvement in hAD symptoms have been seen when using PPARs[209] (one of the candidates in this study), this could benefit cAD. Therefore in the longterm the work presented in this thesis may provide an important starting point for improved treatment and management of cAD.

In conclusion, several genetic associations have been found in cAD, demonstrating that it is a complex disease with a significant genetic component. This validates the hypothesis of the study. Nevertheless, more research is required to identify other associated genes and to further confirm and identify the roles these genes have in the pathogenesis of cAD.

References

1. Akdis CA, Akdis M (2003), Immunological differences between intrinsic and extrinsic types of atopic dermatitis, *Clinical & Experimental Allergy* 33: 1618-1621

2. Olivry T, DeBoer DJ, Prelaud P, Bensignor E (2007), Food for thought: pondering the relationship between canine atopic dermatitis and cutaneous adverse food reactions, *Vet Dermatol* 18: 390-391

3. Olivry T, Hill PB (2001), The ACVD task force on canine atopic dermatitis (VIII): is the epidermal lipid barrier defective?, *Veterinary Immunology and Immunopathology* 81: 215-218

4. Vickery BP (2007), Skin barrier function in atopic dermatitis., *Current Opinion in Pediatrics* 19: 89-93

5. Williams HC (1992), Is the prevalence of atopic dermatitis increasing?, *Clinical and experimental dermatology* 17: 385-391

6. Hillier A, Griffin CE (2001), The ACVD task force on canine atopic dermatitis (I): incidence and prevalence, *Veterinary Immunology and Immunopathology* 81: 147-151

7. Griffin CE, DeBoer DJ (2001), The ACVD task force on canine atopic dermatitis (XIV): clinical manifestations of canine atopic dermatitis., *Veterinary Immunology and Immunopathology* 81: 264

8. Nesbitt GH, Kedan GS, Caciolo P. (1984) Canine atopy, part I. Etiology and diagnosis. *Comp.Cont.Educ.Pract.Vet.* 6, 73-84. 1984.

9. Scott DW, Miller WH, Griffin C.Skin immune system and allergic skin disease. In: Muller and Kirk eds. *Small Animal Dermatology 6th ed.* 6 ed. 2001;543-666.

10. Saridomichelakis MN, Koutinas AF, Gioulekas D, Leontidis L (1999). Canine atopic dermatitis in Greece: clinical observations and the prevalence of positive intradermal test reactions in 91 spontaneous cases. Veterinary Immunology and Immunopathology 69, 61-73. 1999.

11. Reedy LM, Miller Jr. WH, Willemse T*Allergic Skin Diseases of Dogs and Cats*, Saunders, Philadelphia, 1997.

12. Fazakerley J, Nuttall T, Sales D, Schmidt V, Carter SD, Hart CA, McEwan NA (2009), Staphylococcal colonization of mucosal and lesional skin sites in atopic and healthy dogs, *Veterinary Dermatology* 3, 179-84

13. Bloom P (2006), Atopic dermatitis in dogs - Hitting the moving target, *The Veterinary Journal* 171: 16-17

14. Schnabl B, Bettenay SV, Dow K, Mueller RS (2006), Results of allergen-specific immunotherapy in 117 dogs with atopic dermatitis, *The Veterinary Record* 158: 81

15. Olivry T, Mueller RS (2003), Evidence-based veterinary dermatology: a systematic review of the pharmacotherapy of canine atopic dermatitis, *Veterinary dermatology* 14: 121-146

16. Iwasaki T, Hasegawa A (2006), A randomized comparative clinical trial of recombinant canine interferon-gamma (KT-100) in atopic dogs using antihistamine as control, *Veterinary dermatology* 17: 195-200

17. Steffan J, Favrot C, Mueller F (2006), A systematic review and meta-analysis of the efficacy and safety of cyclosporin for the treatment of atopic dermatitis in dogs, *Veterinary dermatology* 17: 3-16

18. Saevik Bente K, Bergvall K, Holm BR, Saijonmaa-Koulumies LE, HEDHAMMAR AKE, Larsen S, Kristensen F (2004), A randomized, controlled study to evaluate the steroid sparing effect of essential fatty acid supplementation in the treatment of canine atopic dermatitis, *Veterinary dermatology* 15: 137-145

19. Holgate ST (1999), The epidemic of allergy and asthma, *Nature* 402: B2-B4

20. Kay AB (2001), Allergy and Allergic Diseases- First of Two Parts, *The New England Journal of Medicine* 344: 30-37

21. Scott D.W. (1981) Observations on canine atopy. Journal of the American Animal Hospital Association 17, 91-100. 1981.

22. Olivry T, Hill PB (2001), The ACVD task force on canine atopic dermatitis (XVIII): histopathology of skin lesions, *Veterinary Immunology and Immunopathology* 81: 305-309

23. Olivry T, Naydan DK, Moore PF (1997), Characterization of the cutaneous inflammatory infiltrate in canine atopic dermatitis, *American Journal of Dermatopathology* 19: 477-486

24. Nimmo Wilkie JS, Yager JA, Wilkie BN, Parker WM (1991), Abnormal cutaneous response to mitogens and a contact allergen in dogs with atopic dermatitis, *Veterinary Immunology and Immunopathology* 28: 97-106

25. Kapp A (1995), A topic dermatitis the skin manifestation of atopic, *Clinical & Experimental Allergy* 25: 210-219

26. Grewe M, Bruijnzeel-Koomen CAFM, Schopf E, Thepen T, Langeveld-Wildschut AG, Ruzicka T, Krutmann J (1998), A role for Th1 and Th2 cells in the immunopathogenesis of atopic dermatitis, *Immunology Today* 19: 359-361

27. Grewe M, Gyufko K, Schopf E, Krutmann J (1994), Lesional expression of interferon-gamma in atopic eczema., *Lancet* 343: 25-26

28. Olivry T, Hill PB (2001), The ACVD task force on canine atopic dermatitis (IX): the controversy surrounding the route of allergen challenge in canine atopic dermatitis., *Veterinary Immunology and Immunopathology* 81: 219-225

29. Owen M.Antigen recognition. In: Roitt I, Brostoff J, and Male D eds. *Immunology*. 5th ed. 1998;14-30.

30. Leung DY (2000), Atopic dermatitis: New insights and opportunities for therapeutic intervention., *J Allergy Clin Immunol* 105: 860-876

31. Nuttall TJ, Knight PA, McAleese SM (2002), Expression of T-helper 1, T-helper 2 and immunosuppressive cytokines in canine atopic dermatitis., *Clinical Experimental Allergy* 32: 789-795

32. Olivry T, Moore PF, Affolter VK (1996), Langerhans cell hyperplasia and IgE expression in canine atopic dermatitis., *Archives of Dermatological Research* 288: 579-585

33. Hill PB, Hillier A, Olivry T (2001), The ACVD task force on canine atopic dermatitis (VI): IgE-induced immediate and late-phase reactions, two inflammatory sequences at sites of intradermal allergen injections., *Veterinary Immunology and Immunopathology* 81: 199-204

34. Halliwell REW, DeBoer DJ (2001), The ACVD task force on canine atopic dermatitis (III): the role of antibodies in canine atopic dermatitis, *Veterinary Immunology and Immunopathology* 81: 159-167

35. Leung DY, Bieber T. (2003), Atopic dermatitis., *Lancet* 361(9352):151-160

36. Akdis CA, Akdis M, Simon D (1999), T cells and T cell-derived cytokines as pathogenic factors in the nonallergic form of atopic dermatitis., *Journal of Investigative Dermatology* 113: 634

37. Cookson W (2004), The immunogenetics of asthma and eczema: a new focus on the epithelium, *Nat Rev Immunol* 4: 978-988

38. Hoffjan S, Epplen JrT (2005), The genetics of atopic dermatitis: recent findings and future options, *Journal of Molecular Medicine* 83: 682-692

39. Galli SJ, Tsai M, Piliponsky AM (2008), The development of allergic inflammation, *Nature* 454: 445-454

40. Sandilands A, Smith FJ, Irvine AD, McLean WH (2007), Filaggrin's fuller figure: a glimpse into the genetic architecture of atopic dermatitis, *J Invest Dermatol* 127: 1282-1284

41. Schleimer RP, Kato A, Kern R, Kuperman D, Avila PC (2007), Epithelium: At the interface of innate and adaptive immune responses, *Journal of Allergy and Clinical Immunology* 120: 1279-1284

42. Rawlings AV, Harding CR (2004), Moisturization and skin barrier function, *Dermatological Therapy* 17 Suppl 1: 43-48

43. Morar N, Willis-Owen SAG, Moffatt MF, Cookson WOCM (2006), The genetics of atopic dermatitis, *Journal of Allergy and Clinical Immunology* 118: 24-34

44. Callard RE, Harper JI (2007), The skin barrier, atopic dermatitis and allergy: a role for Langerhans cells?, *Trends in Immunology* 28: 294-298

45. Proksch E, Folster-Holst R, Jensen JM (2006), Skin barrier function, epidermal proliferation and differentiation in eczema, *Journal of Dermatological Science* 43: 159-169

46. Ohnishi Y, Okino N, Ito M, Imayama S (1999), Ceramidase Activity in Bacterial Skin Flora as a Possible Cause of Ceramide Deficiency in Atopic Dermatitis, *Clinical and Vaccine Immunology* 6: 101-104

47. Nilsson EJ, Henning CG, Magnusson J (1992), Topical corticosteroids and Staphylococcus aureus in atopic dermatitis, *Journal of the American Academy of Dermatology* 27: 29-34

48. Ring J, Abeck D, Neuber K (1992), Atopic eczema: role of microorganisms on the skin surface, *Allergy* 47: 265-269

49. Hara J, Higuchi K, Okamoto R, Kawashima M, Imokawa G (2000), High-Expression of Sphingomyelin Deacylase is an Important Determinant of Ceramide Deficiency Leading to Barrier Disruption in Atopic Dermatitis1, *Journal of Investigative Dermatology* 115: 406-413

50. Kim HT, Lee JY, Han BG, Kimm K, Oh B, Shin HD, Namkung JH, Kim E, Park T, Yang JM (2007), Association analysis of sphingomyelinase 2 polymorphisms for the extrinsic type of atopic dermatitis in Koreans, *Journal of Dermatological Science* 46: 143-146

51. Smith FJ, Irvine AD, Terron-Kwiatkowski A, Sandilands A, Campbell LE, Zhao Y, Liao H, Evans AT, Goudie DR, Lewis-Jones S, Arseculeratne G, Munro CS, Sergeant A, O'Regan G, Bale SJ, Compton JG, DiGiovanna JJ, Presland RB, Fleckman P, McLean WH (2006), Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris, *Nature Genetics* 38: 337-342

52. Palmer CN, Irvine AD, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee SP, Goudie DR, Sandilands A, Campbell LE, Smith FJ, O'Regan GM, Watson RM, Cecil JE, Bale SJ, Compton JG, DiGiovanna JJ, Fleckman P, Lewis-Jones S, Arseculeratne G, Sergeant A, Munro CS, El HB, McElreavey K, Halkjaer LB, Bisgaard H, Mukhopadhyay S, McLean WH (2006), Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis, *Nature Genetics* 38: 441-446

53. Rawlings AV, Matts PJ (2005), Stratum Corneum Moisturization at the Molecular Level : An Update in Relation to the Dry Skin Cycle, *Journal of Investigative Dermatology* 124: 1099-1110

54. Cookson WO, Ubhi B, Lawrence R, Abecasis GR, Walley AJ, Cox HE, Coleman R, Leaves NI, Trembath RC, Moffatt MF, Harper JI (2001), Genetic linkage of

childhood atopic dermatitis to psoriasis susceptibility loci, *Nature Genetics* 27: 372-373

55. Jeong SK, Kim HJ, Youm JK, Ahn SK, Choi EH, Sohn MH, Kim KE, Hong JH, Shin DM, Lee SH (2008), Mite and cockroach allergens activate protease-activated receptor 2 and delay epidermal permeability barrier recovery, *Journal of Investigative Dermatology* 128: 1930-1939

56. Hammad H, Lambrecht BN (2008), Dendritic cells and epithelial cells: linking innate and adaptive immunity in asthma, *Nature Reviews:Immunology* 8: 193-204

57. Sokol CL, Barton GM, Farr AG, Medzhitov R (2008), A mechanism for the initiation of allergen-induced T helper type 2 responses, *Nature Immunology* 9: 310-318

58. Chesney CJ (1995), Measurement of skin hydration in normal dogs and in dogs with atopy or a scaling dermatosis, *Journal of Small Animal Practice* 36: 305-309

59. van de BJ, Simpson JW. (1990) Fat absorption in dogs with atopic dermatitis. In: *Advances in Veterinary Dermatology* (1 edtion) (Von Tscharner C and Halliwell REW eds);pp. 155-160, Bailliere-Tindall

60. Horrobin D, Dobbin S, Reynolds B. (1990) Plasma fatty acids in dogs and their response to essential fatty acid supplementation. In: *Advances in Veterinary Dermatology* (1 edtion) (Von Tscharner C and Halliwell REW eds);pp. 155-160, Bailliere-Tindall

61. Mueller RS, Fettman MJ, Richardson K, Hansen RA, Miller A, Magowitz J, Ogilvie GK (2005), Plasma and skin concentrations of polyunsaturated fatty acids before and after supplementation with n-3 fatty acids in dogs with atopic dermatitis, *American Journal of Veterinary Research* 66: 868-873

62. Inman AO, Olivry T, Dunston SM, Monteiro-Riviere NA, Gatto H (2001), Electron Microscopic Observations of Stratum Corneum Intercellular Lipids in Normal and Atopic Dogs, *Veterinary Pathology* 38: 720-723

63. Marsella R, Girolomoni G (2009), Canine Models of Atopic Dermatitis: A Useful Tool with Untapped Potential, *Journal of Investigative Dermatology* 129(10):2351-7

64. Janeway CA, Jr., Medzhitov R (2002), Innate immune recognition, *Annual Reviews of Immunology* 20: 197-216

65. Novak N (2009), New insights into the mechanism and management of allergic diseases: atopic dermatitis, *Allergy* 64: 265-275

66. Eisenbarth SC, Cassel S, Bottomly K (2004), Understanding asthma pathogenesis: linking innate and adaptive immunity, *Current Opinion in Pediatrics* 16: 659-666

67. De Benedetto A, Agnihothri R, McGirt LY, Bankova LG, Beck LA (2009), Atopic Dermatitis: A Disease Caused by Innate Immune Defects?, *Journal of Investigative Dermatology* 129: 14-30

68. Esche C, de BA, Beck LA (2004), Keratinocytes in atopic dermatitis: inflammatory signals, *Current Allergy and Asthma Reports* 4: 276-284

69. Eder W, Klimecki W, Yu L, von ME, Riedler J, Braun-Fahrlander C, Nowak D, Holst O, Martinez FD (2006), Association between exposure to farming, allergies and genetic variation in CARD4/NOD1, *Allergy* 61: 1117-1124

70. Ahmad-Nejad P, Mrabet-Dahbi S, Breuer K, Klotz M, Werfel T, Herz U, Heeg K, Neumaier M, Renz H (2004), The Toll-like receptor 2 R753Q polymorphism defines a subgroup of patients with atopic dermatitis having severe phenotype, *Journal of Allergy and Clinical Immunology* 113: 565-567

71. Akdis CA, Akdis M, Bieber T, Bindslev-Jensen C, Boguniewicz M, Eigenmann P, Hamid Q, Kapp A, Leung DY, Lipozencic J, Luger TA, Muraro A, Novak N, Platts-Mills TA, Rosenwasser L, Scheynius A, Simons FE, Spergel J, Turjanmaa K, Wahn U, Weidinger S, Werfel T, Zuberbier T (2006), Diagnosis and treatment of atopic dermatitis in children and adults: European Academy of Allergology and Clinical Immunology/American Academy of Allergy, Asthma and Immunology/PRACTALL Consensus Report, *Journal of Allergy and Clinical Immunology* 118: 152-169

72. Hoffjan S, Stemmler S, Parwez Q, Petrasch-Parwez E, Arinir U, Rohde G, Reinitz-Rademacher K, Schultze-Werninghaus G, Bufe A, Epplen JT (2005), Evaluation of the toll-like receptor 6 Ser249Pro polymorphism in patients with asthma, atopic dermatitis and chronic obstructive pulmonary disease, *BMC Medical Genetics* 6: 34

73. Weidinger S, Novak N, Klopp N, Baurecht H, Wagenpfeil S, Rummler L, Ring J, Behrendt H, Illig T (2006), Lack of association between Toll-like receptor 2 and Toll-like receptor 4 polymorphisms and atopic eczema, *Journal of Allergy and Clinical Immunology* 118: 277-279

74. Novak N, Yu CF, Bussmann C, Maintz L, Peng WM, Hart J, Hagemann T, az-Lacava A, Baurecht HJ, Klopp N, Wagenpfeil S, Behrendt H, Bieber T, Ring J, Illig T, Weidinger S (2007), Putative association of a TLR9 promoter polymorphism with atopic eczema, *Allergy* 62: 766-772

75. Weidinger S, Klopp N, Rummler L, Wagenpfeil S, Novak N, Baurecht HJ, Groer W, Darsow U, Heinrich J, Gauger A (2005), Association of NOD1 polymorphisms with atopic eczema and related phenotypes, *Journal of Allergy and Clinical Immunology* 116: 177-184

76. Toke O (2005), Antimicrobial peptides: new candidates in the fight against bacterial infections, *Biopolymers* 80: 717-735

77. Dhople V, Krukemeyer A, Ramamoorthy A (2006), The human beta-defensin-3, an antibacterial peptide with multiple biological functions, *Biochimica et Biophysica Acta* 1758: 1499-1512

78. Hirsch T, Jacobsen F, Steinau HU, Steinstraesser L (2008), Host defense peptides and the new line of defence against multiresistant infections, *Protein and Peptide Letters* 15: 238-243

79. Gambichler T, Skrygan M, Tomi NS, Othlinghaus N, Brockmeyer NH, Altmeyer P, Kreuter A (2008), Differential mRNA expression of antimicrobial peptides and proteins in atopic dermatitis as compared to psoriasis vulgaris and healthy skin, *Int Archives of Allergy and Immunology* 147: 17-24

80. Howell MD (2007), The role of human beta defensins and cathelicidins in atopic dermatitis, *Current Opinion in Allergy & Clinical Immunology.* 7: 413-417

81. van Damme CM, Willemse T, van Dijk A, Haagsman HP, Veldhuizen EJ (2009), Altered cutaneous expression of [beta]-defensins in dogs with atopic dermatitis, *Molecular Immunology* 46: 2449-2455

82. DeBoer DJ, Marsella R (2001), The ACVD task force on canine atopic dermatitis (XII): the relationship of cutaneous infections to the pathogenesis and clinical course of canine atopic dermatitis, *Veterinary Immunology and Immunopathology* 81: 239-249

83. Ferens W.A., Goff W.L., vis W.C., Fox L.K., Deobald C, Hamilton M.J., Hach GA (1998), Induction of type 2 cytokines by a staphylococcal enterotoxin superantigen, *Journal of Natural Toxins.* 7: 193-213

84. MacEwan NA. (2000) Adherence by Staphylococcus intermedius to canine keratinocytes in atopic dermatitis. *Research in Veterinary Science* 68, 279-283

85. Wilson BB, Deuell B, Mills TA. (1993) Atopic dermatitis associated with dermatophyte infection and Trichophyton hypersensitivity. Cutis 51, 191-192

86. Kroger S, Neuber K, Gruseck E, Ring J, Abeck D. (1995) *Pityrosporum ovale* extracts increase interleukin-4, interleukin-10 and IgE synthesis in patients with atopic eczema. Acta Dermato-Venereologica 75(5), 357-360

87. Tengvall LM, Johansson C, Zargari A, Bengtsson A, van der Ploeg I, Jones I, Hardfast B, Scheynius A. (1996) Detection of *Pityrosporum obiculare* reactive T-cells from skin and blood in atopic dermatitis and characterisation of their cytokine profiles. *Clinical Experimental Allergy* 26, 1286-1297

88. Nuttall TJ, Halliwell RE (2001), Serum antibodies to Malassezia yeasts in canine atopic dermatitis, *Veterinary Dermatology* 12: 327-332

89. Farver K, Morris DO, Shofer F, Esch B (2005), Humoral measurement of type-1 hypersensitivity reactions to a commercial Malassezia allergen, *Veterinary Dermatology* 16: 261-268

90. Morris DO, DeBoer DJ (2003), Evaluation of serum obtained from atopic dogs with dermatitis attributable to Malassezia pachydermatis for passive transfer of immediate hypersensitivity to that organism, *American Journal of Veterinary Research* 64: 262-266

91. Morris DO, Clayton DJ, Drobatz KJ, Felsburg PJ (2002), Response to Malassezia pachydermatis by peripheral blood mononuclear cells from clinically normal and atopic dogs, *American Journal of Veterinary Research* 63: 358-362

92. Maintz L, Novak N (2007), Getting more and more complex: the pathophysiology of atopic eczema, *European Journal of Dermatology* 17: 267-283

93. Yazdanbakhsh M, Remsner P.G., van RR (2002), Allergy, parasites, and the hygiene hypothesis, *Science* 296: 490-494

94. Fallon PG, Mangan NE (2007), Suppression of TH2-type allergic reactions by helminth infection, *Nature Reviews Immunology* 7: 220-230

95. Hawrylowicz CM, O'Garra A (2005), Potential role of interleukin-10-secreting regulatory T cells in allergy and asthma, *Nature Reviews Immunology* 5: 271-283

96. Helmer M, Epe C & Mueller, RS (2008), The effect of helminth administration on canine atopic dermatitis- a pilot study, Scientific Abstracts of the Sixth World Congress of Veterinary Dermatology, 19-22 November 2008, Hong Kong. *Veterinary Dermatology* 19(s1): 33.

97. Patki A (2007), Eat dirt and avoid atopy: The hygiene hypothesis revisited, *Indian Journal of Dermatology, Venereology and Leprology* 73: 2-4

98. Schafer T, Merkl J, Klemm E, Wichmann HE, Ring J (2008), We and our pets: allergic together?, *Acta Veterinaria Hungarica.* 56: 153-161

99. Schick RO, Fadok VA (1986), Responses of atopic dogs to regional allergens: 268 cases (1981-1984), *Journal of the American Veterinary Medical Association* 189: 1493-1496

100. Vollset I (1985), Atopic dermatitis in Norwegian dogs, *Nordisk Veterinaer Medicin* 37: 97-106

101. Sousa CA, Marsella R (2001), The ACVD task force on canine atopic dermatitis (II): genetic factors, *Veterinary Immunology and Immunopathology* 81: 153-157

102. Tarpataki N, Papa K, Reiczigel J, Vajdovich P, Vorosi K (2006), Prevalence and features of canine atopic dermatitis in Hungary, *Acta veterinaria Hungarica* 54: 353-366

103. Jaeger K, Linek M, Power HT, Bettenay SV, Zabel S, Rosychuk RA, Mueller RS (2010), Breed and site predispositions of dogs with atopic dermatitis: a comparison of five locations in three continents, *Veterinary dermatology* 21: 118-122

104. Shaw SC, Wood JLN, Freeman J, Littlewood JD, Hannant D (2004), Estimation of heritability of atopic dermatitis in Labrador and Golden Retrievers, *American Journal of Veterinary Research* 65: 1014-1020

105. Larsen F.S., Holm NV, Henningsen K (1986), Atopic dermatitis. A genetic epidemiologic study in a population based twin sample, *Journal of the American Academy of Dermatology* 15: 487-494

106. Schultz LF (1993), Atopic dermatitis: a genetic-epidemiologic study in a population based twin sample, *Journal of the American Academy of Dermatology* 28: 719-723

107. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, Ostrander EA (2004), Extensive and breed-specific linkage disequilibrium in Canis familiaris, *Genome Research* 14: 2388-2396

108. Starkey MP, Scase TJ, Mellersh CS, Murphy S (2005), Dogs really are man's best friend--canine genomics has applications in veterinary and human medicine!, *Briefings in Functional Genomics and Proteomics* 4: 112-128

109. Ubbink GJ, van de BJ, Hazewinkel HA, Rothuizen J (1998), Cluster analysis of the genetic heterogeneity and disease distributions in purebred dog populations, *Veterinary Record* 142: 209-213

110. Risch N, Merikangas K (1996), The future of genetic studies of complex human diseases, *Science* 273: 1516-1517

111. Patterson DF (2000), Companion animal medicine in the age of medical genetics, *Journal of Veterinary Internal Medicine* 14: 1-9

112. Young R.P., Sharp P.A., Lynch J.R., Faux J.A., Lathrop G.M., Cookson W.O., Hopkin J.M. (1992), Confirmation of genetic linkage between atopic IgE responses and chromosome 11q13., *Journal of Medical Genetics* 29: 236-238

113. Beyer K, Nickel R, Freidhoff L, Bjorksten B, Huang SK, Barnes KC, MacDonald S, Forster J, Zepp F, Wahn V, Beaty TH, Marsh DG, Wahn U (2000), Association and Linkage of Atopic Dermatitis with Chromosome 13q12-14 and 5q31-33 Markers, *Journal of Investigative Dermatology* 115: 906-908

114. Lee YA, Wahn U, Kehrt R, Tarani L, Businco L, Gustafsson D, Andersson F, Oranje AP, Wolkertstorfer A, Berg A, Hoffmann U, Kuster W, Wienker T, Ruschendorf F, Reis A (2000), A major susceptibility locus for atopic dermatitis maps to chromosome 3q21, *Nature Genetics* 26: 470-473

115. Bradley M, Soderhall C, Luthman H, Wahlgren CF, Kockum I, Nordenskjold M (2002), Susceptibility loci for atopic dermatitis on chromosomes 3, 13, 15, 17 and 18 in a Swedish population, *Human Molecular Genetics* 11: 1539-1548

116. Haagerup A, Bjerke T, tz PO, Dahl R, Binderup HG, Tan Q, Kruse TA (2004), Atopic Dermatitis: a Total Genome-scan for Susceptibility Genes, *Acta Dermato-Venereologica* 84: 346-352

117. Kurz T, Altmueller J, Strauch K, Ruschendorf F, Heinzmann A, Moffatt MF, Cookson WOCM, Inacio F, Nurnberg P, Stassen HH, Deichmann KA (2005), A genome-wide screen on the genetics of atopy in a multiethnic European population reveals a major atopy locus on chromosome 3q21.3, *Allergy* 60: 192-199

118. Jones G, Wu S, Jang N, Fulcher P, Hogan P, Stewart G (2006), Polymorphisms with the CTLA4 gene are associated with infant atopic dermatitis., *British Journal of Dermatology* 154: 467-471

119. Moffatt MF, Allen MH, Reynolds NJ3, Meggitt ST, Cookson WO, Barker JN (2005), 35th Annual European Society for Dermatological Research (ESDR) Meeting 22-24 September 2005, Tübingen, Germany: A Case-Control Study Investigating Candidate Genes in Adult Eczema, *Journal of Investigative Dermatology* 125: A22

120. Nishio Y, Noguchi E, Ito S, Ichikawa E, Umebayashi Y, Otsuka F, Arinami T (2001), Mutation and association analysis of the interferon regulatory factor 2 gene (IRF2) with atopic dermatitis, *Journal of Human Genetics* 46: 664-667

121. Rafatpanah H, Bennett E, Pravica V, McCoy MJ, David TJ, Hutchinson IV, Arkwright PD (2003), Association between novel GM-CSF gene polymorphisms and the frequency and severity of atopic dermatitis, *Journal of Allergy and Clinical Immunology* 112: 593-598

122. Tsunemi Y, Saeki H, Nakamura K, Sekiya T, Hirai K, Kakinuma T, Fujita H, Asano N, Tanida Y, Wakugawa M, Torii H, Tamaki K (2002), Interleukin-13 gene polymorphism G4257A is associated with atopic dermatitis in Japanese patients, *Journal of Dermatological Science* 30: 100-107

123. Liu X, Beaty TH, Deindl P, Huang SK, Lau S, Sommerfeld C, Fallin MD, Kao WL, Wahn U, Nickel R (2003), Associations between total serum IgE levels and the 6 potentially functional variants within the genes IL4, IL13, and IL4RA in German children: The German Multicenter Atopy Study, *Journal of Allergy and Clinical Immunology* 112: 382-388

124. Kawashima T, Noguchi E, Arinami T, Yamakawa-Kobayashi K, Nakagawa H, Otsuka F, Hamaguchi H (1998), Linkage and association of an interleukin 4 gene polymorphism with atopic dermatitis in Japanese families, *Journal of Medical Genetics* 35: 502-504

125. Walley AJ, Chavanas S, Moffatt MF, Esnouf RM, Ubhi B, Lawrence R, Wong K, Abecasis GR, Jones EY, Harper JI, Hovnanian A, Cookson WOCM (2001), Gene polymorphism in Netherton and common atopic disease, *Nature Genetics* 29: 175-178

126. Nishio Y, Noguchi E, Shibasaki M, Kamioka M, Ichikawa E, Ichikawa K, Unebayashi Y, Otsuka F, Arinami T (2003), Association between polymorphisms in the SPINK5 gene and atopic dermatitis in the Japanese, *Genes and Immunity* 4: 515-517

127. Kato A, Fukai K, Oiso N, Hosomi N, Murakami T, Ishii M (2003), Association of SPINK5 gene polymorphisms with atopic dermatitis in the Japanese population, *British Journal of Dermatology* 148: 665-669

128. Cox, Moffatt, Faux, Walley, Coleman, Trembath, Cookson, Harper (1998), Association of atopic dermatitis to the beta subunit of the high affinity immunoglobulin E receptor, *British Journal of Dermatology* 138: 182-187

129. Novak N, Kruse S, Potreck J, Maintz L, Jenneck C, Weidinger S, Fimmers R, Bieber T (2005), Single nucleotide polymorphisms of the IL18 gene are associated with atopic eczema, *Journal of Allergy and Clinical Immunology* 115: 828-833

130. Chae SC, Song JH, Lee YC, Kim JW, Chung HT (2003), The association of the exon 4 variations of Tim-1 gene with allergic diseases in a Korean population, *Biochemical and Biophysical Research Communications* 312: 346-350

131. Jang N, Stewart G, Jones G (2005), Polymorphisms within the PHF11 gene at chromosome 13q14 are associated with childhood atopic dermatitis, *Genes Immun* 6: 262-264

132. Mao XQ, Shirakawa T, Yoshikawa T, Yoshikawa K, Kawai M, Sasaki S, Enomoto T, Hashimoto T, Furuyama J, Hopkin JM, Morimoto K (1996), Association between genetic variants of mast-cell chymase and eczema, *The Lancet* 348: 581-583

133. Mao XQ, Shirakawa T, Enomoto T, Shimazu S, Dake Y, Kitano H, Hagihara A, Hopkin JM (1998), Association between Variants of Mast Cell Chymase Gene and Serum IgE Levels in Eczema, *Human Heredity* 48: 38-41

134. Tanaka K, Sugiura H, Uehara M, Sato H, Hashimato-Tamaoki T, Furuyama J (1999), Association between mast cell chymase genotype and atopic eczema: comparison between patients with atopic eczema alone and those with atopic eczema and atopic respiratory disease, *Clinical & Experimental Allergy* 29: 800-803

135. Iwanaga T, McEuen A, Walls AF, Clough JB, Keith TP, Rorke S, Barton SJ, Holgate ST, Holloway JW (2004), Polymorphism of the mast cell chymase gene (CMA1) promoter region: lack of association with asthma but association with serum total immunoglobulin E levels in adult atopic dermatitis, *Clinical & Experimental Allergy* 34: 1037-1042

136. Weidinger S, Rummler L, Klopp N, Wagenpfeil S, Baurecht HJ, Fischer G, Holle R, Gauger A, Schafer T, Jakob T, Ollert M, Behrendt H, Wichmann HE, Ring J, Illig T (2005), Association study of mast cell chymase polymorphisms with atopy, *Allergy* 60: 1256-1261

137. Oiso N, Fukai K, Ishii M (2000), Interleukin 4 receptor alpha chain polymorphism Gln551Arg is associated with adult atopic dermatitis in Japan, *British Journal of Dermatology* 142: 1003-1006

138. Hosomi N, Fukai K, Oiso N, Kato A, Ishii M, Kunimoto H, Nakajima K (2004), Polymorphisms in the Promoter of the Interleukin-4 Receptor [alpha] Chain Gene Are Associated with Atopic Dermatitis in Japan, *J Investig Dermatol* 122: 843-845

139. Kabesch M, Peters W, Carr D, Leupold W, Weiland SK, von Mutius E (2003), Association between polymorphisms in caspase recruitment domain containing protein 15 and allergy in two German populations, *Journal of Allergy and Clinical Immunology* 111: 813-817

140. Nickel RG, Casolaro V, Wahn U, Beyer K, Barnes KC, Plunkett BS, Freidhoff LR, Sengler C, Plitt JR, Schleimer RP, Caraballo L, Naidu RP, Levett PN, Beaty TH, Huang SK (2000), Atopic Dermatitis Is Associated with a Functional Mutation in the Promoter of the C-C Chemokine RANTES, *The Journal of Immunology* 164: 1612-1616

141. Tsunemi Y, Saeki H, Nakamura K, Sekiya T, Hirai K, Fujita H, Asano N, Tanida Y, Kakinuma T, Wakugawa M, Torii H, Tamaki K (2002), Eotaxin gene single nucleotide polymorphisms in the promoter and exon regions are not associated with susceptibility to atopic dermatitis, but two of them in the promoter region are associated with serum IgE levels in patients with atopic dermatitis, *Journal of Dermatological Science* 29: 222-228

142. Arkwright PD, Chase JM, Babbage S, Pravica V, David TJ, Hutchinson IV (2001), Atopic dermatitis is associated with a low-producer transforming growth factor [beta]1 cytokine genotype, *Journal of Allergy and Clinical Immunology* 108: 281-284

143. Vasilopoulos Y, Cork MJ, Murphy R, Williams HC, Robinson DA, Duff GW, Ward SJ, Tazi-Ahnini R (2003), Genetic Association Between an AACC Insertion in the 3[prime]UTR of the Stratum Corneum Chymotryptic Enzyme Gene and Atopic Dermatitis, *Journal of Investigative Dermatology* 123: 62-66

144. Vavilin VA (2003), Interaction of GSTM1, GSTT1, and GSTP1 Genotypes in Determination of Predisposition to Atopic Dermatitis, *Bulletin of Experimental Biology & Medicine* 136: 388-391

145. Gray A, Stewart H, Pravica V, Fryer AA, Lenney W, Hutchinson IV, Arkwright PD (2006), Polymorphisms of the Bcl-2 family member bfl-1 in children with atopic dermatitis, *Pediatric Allergy and Immunology* 17: 578-582

146. Mamta S, Kamiya M, Jyotsna B, Balaram G (2007), Association of a chromosome 1q21 locus in close proximity to a late cornified envelope-like proline-rich 1 (LELP1) gene with total serum IgE levels, *Journal of Human Genetics* 52: 378-383

147. Shimizu etal (2005), Functional SNPs in the distal promoter of the ST2 gene are associated with atopic dermatitis, *Human Molecular Genetics* 14: 2919-2927

148. Hirschhorn JN, Daly MJ (2005), Genome wide association studies for common diseases and complex traits, *Nature Genetics* 6: 95-108

149. Short AD, Kennedy LJ, Barnes A, Fretwell N, Jones C, Thomson W, Ollier WER (2007), Hardy Weinberg Expectations in Canine Breeds: Implications for Genetic Studies, *Journal of Heredity* 98: 445-451

150. Bonferroni CE. "Il calcolo delle assicurazioni su gruppi di teste." *In Studi in Onore del Professore Salvatore Ortu Carboni.* 1935;13-60.

151. Benjamini Y, Hochberg Y (1995), Controlling the false discovery rate: A practical and powerful approach to multiple testing, *Journal of the Royal Statistics Society Serie B* 57: 289-300

152. Hillier A, DeBoer DJ (2001), The ACVD task force on canine atopic dermatitis (XVII): intradermal testing, *Veterinary Immunology and Immunopathology* 81: 289-304

153. Olivry T, Marsella R, Iwasaki T, Mueller R (2007), Validation of CADESI-03, a severity scale for clinical trials enrolling dogs with atopic dermatitis, *Veterinary dermatology* 18: 78-86

154. Clements DN, Wood S, Carter SD, Ollier WE (2007), Assessment of the quality and quantity of genomic DNA recovered from canine blood samples by three different extraction methods, *Research in Veterinary Science 85(1):74-9*

155. Lema C, Kohl-White K, Lewis LR, Dao DD (2006), Optimized pH method for DNA elution from buccal cells collected in Whatman FTA cards, *Genetic Testing* 10: 126-130

156. http://www.ensembl.org/biomart/martview. Ensembl BioMart,

157. www.realsnp.com. Sequenom - real SNP,

158. Shendure J (2008), The beginning of the end for microarrays? *Nature Methods* 5: 585-587

159. Curwen V, Eyras E, Andrews TD, Clarke L, Mongin E, Searle SM, Clamp M (2004), The Ensembl automatic gene annotation system, *Genome Research* 14: 942-950

160. http://www.ensembl.org/index.html. Ensembl. Accessed Oct 14, 2009

161. http://www.ncbi.nlm.nih.gov/pubmed/. Pubmed,

162. http://www.celera.com/. Celera,

163. http://www.ncbi.nlm.nih.gov/RefSeq/. Refseq,

164. http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome. NCBI BLAST,

165. http://lgsun.grc.nia.nih.gov/ANOVA. Nia analysis tool,

166. Giussani P, Maceyka M, Le Stunff H, Mikami A, Lepine S, Wang E, Kelly S, Merrill AH, Jr., Milstien S, Spiegel S (2006), Sphingosine-1-Phosphate Phosphohydrolase Regulates Endoplasmic Reticulum-to-Golgi Trafficking of Ceramide, *Molecular and Cellular Biology* 26: 5055-5069

167. Ahn D-H (2005), TNF-alpha Activates MUC2 Transcription via NF-kappaB but Inhibits via JNK Activation, *Cell Physiology & Biochemistry* 15: 29-40

168. Compton etal (2002), Mapping of the associated phenotype of absent granular layer in ichthyosis vulgaris to the epidermal differentiation complex on chromosome 1, *Experimental Dermatology* 11: 518-526

169. Sugiura H, Ebise H, Tazawa T, Tanaka K, Sugiura Y, Uehara M, Kikuchi K, Kimura T (2005), Large-scale DNA microarray analysis of atopic skin lesions shows overexpression of an epidermal differentiation gene cluster in the alternative pathway and lack of protective gene expression in the cornified envelope, *British Journal of Dermatology* 152: 146-149

170. Moffatt MF (2004), SPINK5: A gene for atopic dermatitis and asthma, *Clinical & Experimental Allergy* 34: 325-327

171. E-Y Seo etal (2006), Identification of target genes of atopic dermatitis by real time PCR., *Journal of Investigative Dermatology* 126: 1187-1189

172. Samuelsson L, Stiller C, Friberg C, Nilsson C, Inerot A, Wahlström J (2004), Association Analysis of Cystatin A and Zinc Finger Protein 148, Two Genes Located at the Psoriasis Susceptibility Locus PSORS5, *Journal of Investigative Dermatology* 122: 1399-1400

173. Seguchi T, Chang-Yi C, Shigeru K, Masae T, Kinue A, Tadashi Tezuka ER (1996), Decreased expression of filaggrin in atopic skin, *Archives of Dermatological Research* V288: 442-446

174. Lee CH, Marekov LN, Kim SY, Brahim JS, Park MH, Steinert PM (2000), Small proline-rich protein 1 is the major component of the cell envelope of normal human oral keratinocytes, *FEBS Letters* 477: 268-272

175. An G, Huang TH, Tesfaigzi J, Garcia-Heras J, Ledbetter DH, Carlson DM, Wu R (1992), An unusual expression of a squamous cell marker, small proline-rich protein gene, in tracheobronchial epithelium: differential regulation and gene mapping, *American Journal of Respiratory Cell and Molecular Biology* 7: 104-111

176. Green KJ and Jones JC. (1996), Desmosomes and hemidesmosomes: structure and function of molecular components, *The FASEB Journal* 10: 871-881

177. Y.Kitajima (2002), Mechanisms of desmosome assembly and disassembly, *Clinical & Experimental Dermatology* 27: 684-690

178. Park YD, Lyou YJ, Lee KJ, Lee DY, Yang JM (2006), Towards profiling the gene expression of fibroblasts from atopic dermatitis patients: human 8K complementary DNA microarray, *Clinical & Experimental Allergy* 36: 649-657

179. Chen X, Bonne S, Hatzfeld M, van Roy F, Green KJ (2002), Protein Binding and Functional Characterization of Plakophilin 2. Evidence for its Diverse roles in desmosomes and beta -catenin signalling, *Journal of Biological Chemistry* 277: 10512-10522

180. Ding Y, Shen S, Lino AC, Curotto de Lafaille MA, Lafaille JJ (2008), Beta-catenin stabilization extends regulatory T cell survival and induces anergy in nonregulatory T cells, *Nature Medicine* 14: 162-169

181. Liu YJ, Soumelis V, Watanabe N, Ito T, Wang YH, Malefyt RdW, Omori M, Zhou B, Ziegler SF (2007), TSLP: An Epithelial Cell Cytokine that Regulates T Cell Differentiation by Conditioning Dendritic Cell Maturation, *Annual Review of Immunology* 25: 193-219

182. Yong-Jun Liu (2006), Thymic stromal lymphopoietin: master switch for allergic inflammation, *The Journal of Experimental Medicine* 203: 269-273

183. Tridandapani S, Wang Y, Marsh CB, Anderson CL (2002), Src Homology 2 Domain-Containing Inositol Polyphosphate Phosphatase Regulates NF-{kappa}B-Mediated Gene Transcription by Phagocytic Fc{gamma}Rs in Human Myeloid Cells, *The Journal of Immunology* 169: 4370-4378

184. Hagihara K, Nishikawa T, Sugamata Y, Song J, Isobe T, Taga T, Yoshizaki K (2005), Essential role of STAT3 in cytokine-driven NF-kappaB-mediated serum amyloid A gene expression, *Genes to Cells* 10: 1051-1063

185. O'Hara R, Murphy EP, Whitehead AS, Fitzgerland O, Bresnihan B (2004), Local expression of the serum amyloid A and formyl peptide receptor-like 1 genes in synovial tissue is associated with matrix metalloproteinase production in patients with inflammatory arthritis, *Arthritis & Rheumatism* 50: 1788-1799

186. Sakata Y, Arima K, Takai T, Sakurai W, Masumoto K, Yuyama N, Suminami Y, Kishi F, Yamashita T, Kato T, Ogawa H, Fujimoto K, Matsuo Y, Sugita Y, Izuhara K (2004), The Squamous Cell Carcinoma Antigen 2 Inhibits the Cysteine Proteinase Activity of a Major Mite Allergen, Der p 1, *Journal of Biological Chemistry* 279: 5081-5087

187. Roth J, Vogl T, Sorg C, Sunderkotter C (2003), Phagocyte-specific S100 proteins: a novel group of proinflammatory molecules, *Trends in Immunology* 24: 155-158

188. Cui X, Rouhani FN, Hawari F, Levine SJ (2003), An Aminopeptidase, ARTS-1, Is Required for Interleukin-6 Receptor Shedding, *Journal of Biological Chemistry* 278: 28677-28685

189. Yangagisawa He (1997), The expression of ST2 gene in helper T cells in the binding of ST2 protien to myeloma-dervived RPMI8226 cells, *The Journal of Immunology* 121: 95-103

190. Bock O (2001), Expression of dipeptidyl-peptidase IV (CD26) on CD8+ T cells is significantly decreased in patients with psoriasis vulgaris and atopic dermatitis, *Experimental Dermatology* 10: 414-419

191. Reinhold D, Biton A, Goihl A, Pieper S, Lendeckel U, Faust J, Neubert K, Bank U, Tager M, Ansorge S, Brocke S (2007), Dual Inhibition of Dipeptidyl Peptidase IV and Aminopeptidase N Suppresses Inflammatory Immune Responses, *Annals of the New York Academy of Sciences* 1110: 402-409

192. Busso N, Wagtmann N, Herling C, Chobaz-Peclat V, Bischof-Delaloye A, So A, Grouzmann E (2005), Circulating CD26 Is Negatively Associated with Inflammation in Human and Experimental Arthritis, *American Journal of Pathology* 166: 433-442

193. Ghersi G, Dong H, Goldstein LA, Yeh Y, Hakkinen L, Larjava HS, Chen WT (2002), Regulation of Fibroblast Migration on Collagenous Matrix by a Cell Surface Peptidase Complex, *Journal of Biological Chemistry* 277: 29231-29241

194. Cilla S, Maria B, Ingrid K, Carl-Fredrik W, Holger L, Magnus N (2001), Linkage and association to candidate regions in Swedish atopic dermatitis families, *Human Genetics* V109: 129-135

195. Takayama G, Arima K, Kanaji T, Toda S, Tanaka H, Shoji S, McKenzie ANJ, Nagai H, Hotokebuchi T, Izuhara K (2006), Periostin: A novel component of subepithelial fibrosis of bronchial asthma downstream of IL-4 and IL-13 signals, *Journal of Allergy and Clinical Immunology* 118: 98-104

196. Katoh N, Hirano S, Sueehiro M, Ikenaga K, Yasuno H (2002), Increased levels of serum tissue inhibitor of metalloproteinase-1 but not metalloproteinase-3 in atopic dermatitis, *Clinical & Experimental Immunology* 127: 283-288

197. Kipreos ET, Lander LE, Wing JP, He WW, Hedgecock EM (1996), cul-1 Is Required for Cell Cycle Exit in C. elegans and Identifies a Novel Gene Family, *Cell* 85: 829-839

198. Day RM, Mitchell TJ, Knight SC, Forbes A (2003), Regulation of epithelial syndecan-1 expression by inflammatory cytokines, *Cytokine* 21: 224-233

199. Ekelund E, Saaf A, Tengvall-Linder M, Melen E, Link J, Barker J, Reynolds NJ, Meggitt SJ, Kere J, Wahlgren CF, Pershagen G, Wickman M, Nordenskjold M, Kockum I, Bradley M (2006), Elevated Expression and Genetic Association Links the *SOCS3* Gene to Atopic Dermatitis, *American Journal of Human Genetics* 78: 1060-1065

200. Goren I, Linke A, Muller E, Pfeilschifter J, Frank S (2005), The Suppressor of Cytokine Signaling-3 Is Upregulated in Impaired Skin Repair: Implications for Keratinocyte Proliferation, *Journal of Investigative Dermatology* 126: 477-485

201. Mukoyama Y, Zhou S, Miyachi Y, Matsuyoshi N (2005), T-Cadherin Negatively Regulates the Proliferation of Cutaneous Squamous Carcinoma Cells, *Journal of Investigative Dermatology* 124: 833-838

202. Mukoyama Y, Utani A, Matsui S, Zhou S, Miyachi Y, Matsuyoshi N (2007), T-cadherin enhances cell-matrix adhesiveness by regulating beta1 integrin trafficking in cutaneous squamous carcinoma cells, *Genes to Cells* 12: 787-796

203. NCBI (2009) Phospholipase http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=89869&ordinalpos=4&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum.

204. Liu Y, Krueger JG, Bowcock AM (2007), Psoriasis: genetic associations and immune system changes, *Genes Immun* 8: 1-12

205. Westergaard M, Henningsen J, Svendsen ML, Johansen C, Jensen UB, Schroder HD, Kratchmarova I, Berge RK, Iversen L, Bolund L, Kragballe K, Kristiansen K (2001), Modulation of Keratinocyte Gene Expression and Differentiation by PPAR-Selective Ligands and Tetradecylthioacetic Acid, *Journal of Investigative Dermatology* 116: 702-712

206. Rivier M, Safonova I, Lebrun P, Griffiths CE, Ailhaud G, Michel S (1998), Differential expression of peroxisome proliferator-activated receptor subtypes during the differentiation of human keratinocytes., *Journal of Investigative Dermatology* 111: 1116-1121

207. Ellis CN, Varani J, Fisher GJ, Zeigler ME, Pershadsingh HA, Benson SC, Chi Y, Kurtz TW (2000), Troglitazone Improves Psoriasis and Normalizes Models of Proliferative Skin Disease: Ligands for Peroxisome Proliferator-Activated Receptor-{gamma} Inhibit Keratinocyte Proliferation, *Archives of Dermatology* 136: 609-616

208. Rotraut M, Rolf K, Philipp M, Ullrich K, GâÂtz AW, Jârgen B-Â, Andreas Z, Christine N, Inke RK, Kristian Reich ER (2004), Variations in the genes encoding the peroxisome proliferator-activated receptors a and ? in psoriasis, *Archives of Dermatological Research* V296: 1-5

209. Sertznig P, Seifert M, Tilgen W, Reichrath J, rg [1] (2008), Peroxisome Proliferator-Activated Receptors (PPARs) and the Human Skin: Importance of PPARs in Skin Physiology and Dermatologic Diseases, *American Journal of Clinical Dermatology* 9: 15-31

210. Benoit S, Toksoy A, Ahlmann M, Schmidt M, Sunderkotter C, Foell D, Pasparakis M, Roth J, Goebeler M (2006), Elevated serum levels of calcium-binding S100 proteins A8 and A9 reflect disease activity and abnormal differentiation of keratinocytes in psoriasis, *British Journal of Dermatology* 155: 62-66

211. Frosch M (2006), Myeliod-related protiens 8 and 14 are specifically secreated during interaction of phagocytes and activated endothelium and are useful markers for monitoring diease activity in pauciaticular-onset juvenile rheumatiod arthritis., *Arthritis & Rheumatism* 43: 628-637

212. Chen W, Khurana Hershey GK (2007), Signal transducer and activator of transcription signals in allergic disease, *Journal of Allergy and Clinical Immunology* 119: 529-541

213. Mudter J, Weigmann B, Bartsch B, Kiesslich R, Strand D, Galle PR, Lehr HA, Schmidt J, Neurath MF (2005), Activation Pattern of Signal Transducers and Activators of Transcription (STAT) Factors in Inflammatory Bowel Diseases, *The American Journal of Gastroenterology* 100: 64-72

214. Eriksen KW, Lovato P, Skov L, Krejsgaard T, Kaltoft K, Geisler C, Odum N (2005), Increased sensitivity to interferon-alpha in psoriatic T cells, *Journal of Investigative Dermatology* 125: 936-944

215. Deepak Sampath, Mario Castro, Dwight C.Look, Michael J.Holtzman (1999), Constitutive activation of an epithelial signal transducer and activator of transcription (STAT) pathway in asthma, *Journal of Clinical Investigation.* 103: 1353-1361

216. Ghaffar, Christodoulopoulos, Lamkhioued, Wright, Ihaku, Nakamura, Frenkiel, Hamid (2000), In vivo expression of signal transducer and activator of transcription factor 6 (STAT6) in nasal mucosa from atopic allergic rhinitis: effect of topical corticosteroids, *Clinical & Experimental Allergy* 30: 86-93

217. Darnell JE, Jr., Kerr IM, Stark GR (1994), Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins, *Science* 264: 1415-1421

218. Sakaguchi S (2004), Naturally Arising CD4+ Regulatory T Cells for Immunologic Self-Tolerance and Negative Control of Immune Responses, *Annual Review of Immunology* 22: 531-562

219. Refaeli Y, Van PL, London CA, Tschopp J, Abbas AK (1998), Biochemical mechanisms of IL-2-regulated Fas-mediated T cell apoptosis, *Immunity* 8: 615-623

220. Lin W, Truong N, Grossman WJ, Haribhai D, Williams CB, Wang J, Martin MG, Chatila TA (2005), Allergic dysregulation and hyperimmunoglobulinemia E in Foxp3 mutant mice, *Journal of Allergy and Clinical Immunology* 116: 1106-1115

221. Elias PM, Goerke J, Friend DS (1977), Mammalian epidermal barrier layer lipids: composition and influence on structure, *Journal of Investigative Dermatology* 69: 535-546

222. Candi E, Schmidt R, Melino G (2005), The cornified envelope: a model of cell death in the skin, *Nature Reviews Molecular Cell Biology* 6: 328-340

223. Marekov LN, Steinert PM (1998), Ceramides Are Bound to Structural Proteins of the Human Foreskin Epidermal Cornified Cell Envelope, *Journal of Biological Chemistry* 273: 17763-17770

224. Reiter LV, Torres SM, Wertz PW (2009), Characterization and quantification of ceramides in the nonlesional skin of canine patients with atopic dermatitis compared with controls, *Veterinary Dermatology* 20: 260-266

225. Trautmann A, Altznauer F, Akdis M, Simon HU, Disch R, Brocker EB, Blaser K, Akdis CA (2001), The Differential Fate of Cadherins during T-Cell-Induced Keratinocyte Apoptosis Leads to Spongiosis in Eczematous Dermatitis, *Journal of Investigative Dermatology* 117: 927-934

226. Leppa S, Mali M, Miettinen H, Jalkanen M (1992), Syndecan expression regulates cell morphology and growth of mouse mammary epithelial tumor cells. *Proceedings of the National Academy of Sciences USA* 89: 932

227. Bayer-Garner IB, Dilday BR, Sanderson RD, Smoller BR (2001), Acantholysis and spongiosis are associated with loss of syndecan-1 expression, *Journal of Cutaneous Pathology* 28: 135-139

228. Morar N, Cookson WOCM, Harper JI, Moffatt MF (2007), Filaggrin Mutations in Children with Severe Atopic Dermatitis, *Journal of Investigative Dermatology* 127: 1667-1672

229. Sandilands A, Terron-Kwiatkowski A, Hull PR, O'Regan GM, Clayton TH, Watson RM, Carrick T, Evans AT, Liao H, Zhao Y, Campbell LE, Schmuth M, Gruber R, Janecke AR, Elias PM, van Steensel MAM, Nagtzaam I, van Geel M, Steijlen PM, Munro CS, Bradley DG, Palmer CNA, Smith FJD, McLean WHI, Irvine AD (2007), Comprehensive analysis of the gene encoding filaggrin uncovers prevalent and rare mutations in ichthyosis vulgaris and atopic eczema, *Nature Genetics* 39: 650-654

230. Weidinger S, Rodriguez E, Stahl C, Wagenpfeil S, Klopp N, Illig T, Novak N (2006), Filaggrin Mutations Strongly Predispose to Early-Onset and Extrinsic Atopic Dermatitis, *Journal of Investigative Dermatology* 127: 724-726

231. Nomura T, Sandilands A, Akiyama M, Liao H, Evans AT, Sakai K, Ota M, Sugiura H, Yamamoto K, Sato H, Palmer CNA, Smith FJD, McLean WHI, Shimizu H (2007), Unique mutations in the filaggrin gene in Japanese patients with ichthyosis vulgaris and atopic dermatitis, *Journal of Allergy and Clinical Immunology* 119: 434-440

232. Reiss U, Oskouian B, Zhou J, Gupta V, Sooriyakumaran P, Kelly S, Wang E, Merrill AH, Jr., Saba JD (2004), Sphingosine-phosphate Lyase Enhances Stress-induced Ceramide Generation and Apoptosis, *Journal of Biological Chemistry* 279: 1281-1290

233. Takahashi H, Komatsu N, Ibe M, Ishida-Yamamoto A, Hashimoto Y, Iizuka H (2007), Cystatin A suppresses ultraviolet B-induced apoptosis of keratinocytes, *Journal of Dermatological Science* 46: 179-187

234. NCBI (2009) Cystatin A http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=1475&ordinalpos=1&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum.

235. Kato T, Takai T, Mitsuishi K, Okumura K, Ogawa H (2005), Cystatin A inhibits IL-8 production by keratinocytes stimulated with Der p 1 and Der f 1: Biochemical skin barrier against mite cysteine proteases, *Journal of Allergy and Clinical Immunology* 116: 169-176

236. Vasilopoulos Y, Cork MJ, Teare D, Marinou I, Ward SJ, Duff GW, Tazi-Ahnini R (2007), A nonsynonymous substitution of cystatin A, a cysteine protease inhibitor of house dust mite protease, leads to decreased mRNA stability and shows a significant association with atopic dermatitis, *Allergy* 62: 514-519

237. Komatsu N, Takata M, Otsuki N, Ohka R, Amano O, Takehara K, Saijoh K (2002), Elevated Stratum Corneum Hydrolytic Activity in Netherton Syndrome Suggests an Inhibitory Regulation of Desquamation by SPINK5-Derived Peptides, 118: 436-443

238. Ong C, O'Toole EA, Ghali L, Malone M, Smith VV, Callard R, Harper JI (2004), LEKTI demonstrable by immunohistochemistry of the skin: a potential diagnostic skin test for Netherton syndrome, *British Journal of Dermatology* 151: 1253-1257

239. Soumelis V, Reche PA, Kanzler H, Yuan W, Edward G, Homey B, Gilliet M, Ho S, Antonenko S, Lauerma A, Smith K, Gorman D, Zurawski S, Abrams J, Menon S, McClanahan T, Waal-Malefyt Rd, Bazan F, Kastelein RA, Liu YJ (2002), Human epithelial cells trigger dendritic cell-mediated allergic inflammation by producing TSLP, *Nat Immunol* 3: 673-680

240. Yoo J, Omori M, Gyarmati D, Zhou B, Aye T, Brewer A, Comeau MR, Campbell DJ, Ziegler SF (2005), Spontaneous atopic dermatitis in mice expressing an inducible thymic stromal lymphopoietin transgene specifically in the skin, *The Journal of Experimental Medicine* 202: 541-549

241. Li M, Messaddeq N, Teletin M, Pasquali JL, Metzger D, Chambon P (2005), Retinoid X receptor ablation in adult mouse keratinocytes generates an atopic dermatitis triggered by thymic stromal lymphopoietin, *Proceedings of the National Academy of Science U.S.A* 102: 14795-14800

242. Ebner S, Nguyen VA, Forstner M, Wang YH, Wolfram D, Liu YJ, Romani N (2007), Thymic stromal lymphopoietin converts human epidermal Langerhans cells into antigen presenting cells that induce pro-allergic T cells, *Journal of Allergy and Clinical Immunology* 119: 982-990

243. Stary G, Bangert C, Stingl G, Kopp T (2005), Dendritic Cells in Atopic Dermatitis: Expression of FcεRI on Two Distinct Inflammation-Associated Subsets, *International Archives of Allergy and Immunology* 138: 278-290

244. Novak N, Bieber T (2005), The role of dendritic cell subtypes in the pathophysiology of atopic dermatitis, *Journal of the American Academy of Dermatology* 53: S171-S176

245. Stingl G, Maurer D (1997), IgE-mediated allergen presentation via Fc epsilon RI on antigen-presenting cells, *International Archives of Allergy and Immunology* 113: 24-29

246. NCBI (2008) IL1RL1 T cell function http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=9173&ordinalpos=1&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum.

247. Reinhold D, Kahne T, Steinbrecher A, Wrenger S, Neubert K, Ansorge S, Brocke S (2002), The role of dipeptidyl peptidase IV (DP IV) enzymatic activity in T cell activation and autoimmunity, *Biological Chemistry* 383: 1133-1138

248. Zhou S, Matsuyoshi N, Liang SB, Takeuchi T, Ohtsuki Y, Miyachi Y (2002), Expression of T-Cadherin in Basal Keratinocytes of Skin, *Journal of Investigative Dermatology* 118: 1080-1084

249. Zhou NM (2003), Reciprocal altered expression of T-cadherin and P-cadherin in psoriasis vulgaris, *British Journal of Dermatology* 149: 268-273

250. Dahten A, Mergemeier S, Worm M (2007), PPARgamma expression profile and its cytokine driven regulation in atopic dermatitis, *Allergy* 62: 926-933

251. Bustin SA, Nolan T (2004), Pitfalls of Quantitative Real-Time Reverse-Transcription Polymerase Chain Reaction, *Journal of Biomolecular Techniques* 15: 155-166

252. Bustin SA (2002), Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems, *Journal of Molecular Endocrinology* 29: 23-39

253. Imbeaud S, Graudens E, Boulanger V, Barlet X, Zaborski P, Eveno E, Mueller O, Schroeder A, Auffray C (2005), Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces, *Nucleic Acids Research* 33: e56

254. Lekanne Deprez RH, Fijnvandraat AC, Ruijter JM, Moorman AF (2002), Sensitivity and accuracy of quantitative real-time polymerase chain reaction using SYBR green I depends on cDNA synthesis conditions, *Analytical Biochemistry* 307: 63-69

255. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T (2006), The RIN: an RNA integrity number for assigning integrity values to RNA measurements, *BMC Molecular Biology* 7: 3

256. Huggett J, Dheda K, Bustin S, Zumla A (2005), Real-time RT-PCR normalisation; strategies and considerations, *Genes and Immunity* 6: 279-284

257. Maccoux L, Clements D, Salway F, Day P (2007), Identification of new reference genes for the normalisation of canine osteoarthritic joint tissue transcripts from microarray data, *BMC Molecular Biology* 8: 62

258. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002), Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, *Genome Biology* 3: research0034

259. Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G, Zumla A (2004), Validation of housekeeping genes for normalizing RNA expression in real-time PCR, *Biotechniques* 37: 112-119

260. Mohammadi M, Day PJR (2004), Oligonucleotides used as template calibrators for general application in quantitative polymerase chain reaction, *Analytical Biochemistry* 335: 299-304

261. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004), Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations, *Biotechnology Letters* 26: 509-515

262. Andersen CL, Jensen JL, Orntoft TF (2004), Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets, *Cancer Research* 64: 5245-5250

263. Clements D, Carter S, Innes J, Ollier W, Day P (2006), Analysis of normal and osteoarthritic canine cartilage mRNA expression by quantitative polymerase chain reaction, *Arthritis Research & Therapy* 8: R158

264. Pfaffl MW (2001), A new mathematical model for relative quantification in real-time RT-PCR, *Nucleic Acids Research* 29: e45

265. Rutledge RG, Cote C (2003), Mathematics of quantitative kinetic PCR and the application of standard curves, *Nucleic Acids Research* 31: e93

266. Brinkhof B, Spee B, Rothuizen J, Penning LC (2006), Development and evaluation of canine reference genes for accurate quantification of gene expression, *Analytical Biochemistry* 356: 36-43

267. Penning LC, Vrieling HE, Brinkhof B, Riemers FM, Rothuizen J, Rutteman GR, Hazewinkel HAW (2007), A validation of 10 feline reference genes for gene expression measurements in snap-frozen tissues, *Veterinary Immunology and Immunopathology* 120: 212-222

268. Peters IR, Peeters D, Helps CR, Day MJ (2007), Development and application of multiple internal reference (housekeeper) gene assays for accurate normalisation of canine gene expression studies, *Veterinary Immunology and Immunopathology* 117: 55-66

269. de Mora F, de la Fuente C, Jasmin P, Gatto H, Marco A, Ferrer L, Fondati A, Fondevila D, Torres R (2007), Evaluation of the expression of P-selectin, ICAM-1,

and TNF-alpha in bacteria-free lesional skin of atopic dogs with low-to-mild inflammation, *Veterinary Immunology and Immunopathology* 115: 223-229

270. Wang HB, Wang JT, Zhang L, Geng ZH, Xu WL, Xu T, Huo Y, Zhu X, Plow EF, Chen M, Geng JG (2007), P-selectin primes leukocyte integrin activation during inflammation, *Nature Immunology* 8: 882-892

271. Schmuth M, Jiang YJ, Dubrac S, Elias PM, Feingold KR (2008), Peroxisome proliferator-activated receptors (PPAR) and liver X receptors (LXR) in epidermal biology, *Journal of Lipid Research* R800001-RJLR200

272. Bonness S, Bieber T (2007), Molecular basis of atopic dermatitis., *Current Opinion in Allergy & Clinical Immunology.* 7: 382-386

273. Thorn ZYL (2004), Regulation of the Human Acute Phase Serum Amyloid A Genes by Tumour Necrosis Factor-α, Interleukin-6 and Glucocorticoids in Hepatic and Epithelial Cell Lines, *Scandinavian Journal of Immunology* 59: 152-158

274. Merryman-Simpson AE, Wood SH, Fretwell N, Jones PG, McLaren WM, McEwan NA, Clements DN, Carter SD, Ollier WE, Nuttall T (2008), Gene (mRNA) expression in canine atopic dermatitis: microarray analysis, *Veterinary dermatology* 19: 59-66

275. Wood SH, Clements DN, McEwan NA, Nuttall T, Carter SD (2008) Reference genes for Canine Skin when using Quantitative Real Time PCR. *Veterinary Immunology and Immunopathology* 126(3-4):392-5

276. Olivry T, Dunston SM, Pluchino K, Porter K, Hammerberg B (2008), Lack of detection of circulating skin-specific IgE autoantibodies in dogs with moderate or severe atopic dermatitis, *Veterinary Immunology and Immunopathology* 122: 182-187

277. Roche Applied Biosciences (2008) Roche Universal probe Library Designer: https://www.roche-applied-science.com/servlet/.

278. Kashiwada M, Cattoretti G, McKeag L, Rouse T, Showalter BM, Al-Alem U, Niki M, Pandolfi PP, Field EH, Rothman PB (2006), Downstream of Tyrosine Kinases-1 and Src Homology 2-Containing Inositol 5'-Phosphatase Are Required for Regulation of CD4+CD25+ T Cell Development, *The Journal of Immunology* 176: 3958-3965

279. Avgerinou G, Goules AV, Stavropoulos PG, Katsambas AD (2008), Atopic dermatitis: new immunologic aspects, *International Journal of Dermatology* 47: 219-224

280. Tartaglia-Polcini A, Bonnart C, Micheloni A, Cianfarani F, Andre A, Zambruno G, Hovnanian A, D'Alessio M (2005), SPINK5, the Defective Gene in Netherton Syndrome, Encodes Multiple LEKTI Isoforms Derived from Alternative Pre-mRNA Processing, *Journal of Investigative Dermatology* 126: 315-324

281. Greenbaum D, Colangelo C, Williams K, Gerstein M (2003), Comparing protein abundance and mRNA expression levels on a genomic scale, *Genome Biology* 4: 117

282. Filipowicz W, Bhattacharyya SN, Sonenberg N (2008), Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Reviews Genetics* 9: 102-114

283. Hla T (2003), Signaling and biological actions of sphingosine 1-phosphate, *Pharmacological Research* 47: 401-407

284. Herzinger T, Kleuser B, Schafer-Korting M, Korting C (2007), Sphingosine-1-Phosphate Signaling and the Skin, *American Journal of Clinical Dermatology* 8: 329-336

285. Takahiro Satoh, Michiyo Kaneko, Ming-Hua Wu, Hiroo Yokozeki, Kiyoshi Nishioka (2002), Contribution of selectin ligands to eosinophil recruitment into the skin of patients with atopic dermatitis, *European journal of immunology* 32: 1274-1281

286. Manolis NS, Rosanna M, Kenneth WL, Robert E.Esch, Rania F, Alexander FK (2008), Assessment of cross-reactivity among five species of house dust and storage mites, *Veterinary dermatology* 19: 67-76

287. Brockow K, Hautmann C, Fotisch K, Rakoski J, Borelli S, Vieths S, Ring J (2003), Orange-induced skin lesions in patients with atopic eczema: evidence for a non-IgE-mediated mechanism, *Acta Dermato-Venereologica* 83: 44-48

288. Nodtvedt A, Bergvall K, Sallander M, Egenvall A, Emanuelson U, Hedhammar A (2007), A case-control study of risk factors for canine atopic dermatitis among boxer, bullterrier and West Highland white terrier dogs in Sweden, *Veterinary dermatology* 18: 309-315

289. Kerschenlohr K, Darsow U, Burgdorf WH, Ring J, Wollenberg A (2004), Lessons from atopy patch testing in atopic dermatitis, *Current Allergy and Asthma Reports* 4: 285-289

290. Fuiano N, Incorvaia C, Prodam F, Procaccini DA, Bona G (2008), Relationship between the atopy patch test and clinical expression of the disease in children with atopic eczema/dermatitis syndrome and respiratory symptoms, *Annuals of Allergy Asthma and Immunology* 101: 174-178

291. Bisgaard H, Simpson A, Palmer CNA, nnelykke K, Mclean I, Mukhopadhyay S, Pipper CB, Halkjaer LB, Lipworth B, Hankinson J, Woodcock A, Custovic A (2008), Gene-Environment Interaction in the Onset of Eczema in Infancy: Filaggrin Loss-of-Function Mutations Enhanced by Neonatal Cat Exposure, *PLoS Medicine* 5: e131

292. Sanger F, Nicklen S, Coulson AR (1977), DNA sequencing with chain-terminating inhibitors, *Proceedings of the National Academy of Science U.S.A* 74: 5463-5467

293. Suh Y, Vijg J (2005), SNP discovery in associating genetic variation with human disease phenotypes, *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 573: 41-53

294. Livingston RJ, von NA, Jegga AG, Crawford DC, Carlson CS, Rieder MJ, Gowrisankar S, Aronow BJ, Weiss RB, Nickerson DA (2004), Pattern of sequence variation across 213 environmental response genes, *Genome Research* 14: 1821-1831

295. Wang H, Chu W, Das SK, Zheng Z, Hasstedt SJ, Elbein SC (2003), Molecular screening and association studies of retinoid-related orphan receptor gamma (RORC): a positional and functional candidate for type 2 diabetes *Molecular Genetic Metabolism* 79: 176-182

296. Sentinelli F, Romeo S, Arca M, Filippi E, Leonetti F, Banchieri M, Di MU, Baroni MG (2002), Human resistin gene, obesity, and type 2 diabetes: mutation analysis and population study, *Diabetes* 51: 860-862

297. Casadei S, Cortesi L, Pensotti V, Radice P, Pierotti M, Amadori D, Calistri D (2001), Detection of germline BRCA1 mutations by Multiple-Dye Cleavase Fragment Length Polymorphism (MD-CFLP) method, *British Journal of Cancer* 85: 845-849

298. Krawczynski MR, Czarny-Ratajczak M, Pecold K, Latos-Bielenska A (2004), Only neutral polymorphisms found in the TIGR/myocilin gene of 45 Polish patients with primary open-angle glaucoma, *Journal of Applied Genetics* 45: 275-279

299. Seo JH, Cho DY, Ahn SH, Yoon KS, Kang CS, Cho HM, Lee HS, Choe JJ, Choi CW, Kim BS, Shin SW, Kim YH, Kim JS, Son GS, Lee JB, Koo BH (2004), BRCA1 and BRCA2 germline mutations in Korean patients with sporadic breast cancer, *Human Mutation* 24: 350

300. Glatt CE, DeYoung JA, Delgado S, Service SK, Giacomini KM, Edwards RH, Risch N, Freimer NB (2001), Screening a large reference sample to identify very low frequency sequence variants: comparisons between two genes, *Nature Genetics* 27: 435-438

301. Lilleberg SL (2003), In-depth mutation and SNP discovery using DHPLC gene scanning, *Current Opinion in Drug Discovery and Development* 6: 237-252

302. Penfornis A, Yan G, Shi L, Faustman DL (2003), Polymorphisms of human TAP2 detected by denaturing gradient gel electrophoresis, *Human Immunology* 64: 156-167

303. Saarikoski ST, Suitiala T, Holmila R, Impivaara O, Jarvisalo J, Hirvonen A, Husgafvel-Pursiainen K (2004), Identification of genetic polymorphisms of CYP2S1 in a Finnish Caucasian population, *Mutation Research* 554: 267-277

304. Lee JH, Choi JH, Namkung W, Hanrahan JW, Chang J, Song SY, Park SW, Kim DS, Yoon JH, Suh Y, Jang IJ, Nam JH, Kim SJ, Cho MO, Lee JE, Kim KH, Lee MG (2003), A haplotype-based molecular analysis of CFTR mutations associated with respiratory and pancreatic diseases, *Human Molecular Genetics* 12: 2321-2332

305. Bounpheng M, McGrath S, Macias D, van Orsouw N, Suh Y, Rines D, Vijg J (2003), Rapid, inexpensive scanning for all possible BRCA1 and BRCA2 gene sequence variants in a single assay: implications for genetic testing, *Journal of Medical Genetics* 40: e33

306. Kuhn DN, Borrone J, Meerow AW, Motamayor JC, Brown JS, Schnell RJ (2005), Single-strand conformation polymorphism analysis of candidate genes for reliable identification of alleles by capillary array electrophoresis, *Electrophoresis* 26: 112-125

307. Rossetti S, Englisch S, Bresin E, Pignatti PF, Turco AE (1997), Detection of mutations in human genes by a new rapid method: cleavage fragment length polymorphism analysis (CFLPA), *Molcular and Cellular Probes* 11: 155-160

308. Ganguly A, Rock MJ, Prockop DJ (1993), Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double-stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes, *Proceedings of the National Academy of Science U.S.A* 90: 10325-10329

309. Fischer SG, Lerman LS (1979), Length-independent separation of DNA restriction fragments in two-dimensional gel electrophoresis, *Cell* 16: 191-200

310. Riesner D, Steger G, Zimmat R, Owens RA, Wagenhofer M, Hillen W, Vollbach S, Henco K (1989), Temperature-gradient gel electrophoresis of nucleic acids: analysis of conformational transitions, sequence variations, and protein-nucleic acid interactions, *Electrophoresis* 10: 377-389

311. Sheffield VC, Cox DR, Lerman LS, Myers RM (1989), Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes, *Proceedings of the National Academy of Science U.S.A* 86: 232-236

312. Vijg J, van Orsouw NJ (1999), Two-dimensional gene scanning: exploring human genetic variability, *Electrophoresis* 20: 1239-1249

313. van Orsouw NJ, Dhanda RK, Elhaji Y, Narod SA, Li FP, Eng C, Vijg J (1999), A highly accurate, low cost test for BRCA1 mutations, *Journal of Medical Genetics* 36: 747-753

314. McGrath SB, Bounpheng M, Torres L, Calavetta M, Scott CB, Suh Y, Rines D, van ON, Vijg J (2001), High-speed, multicolor fluorescent two-dimensional gene scanning, *Genomics* 78: 83-90

315. Oefner P, Underhill PA (1995) Comparative DNA sequencing by denaturing high-performance liquid chromatography. *American Journal of Human Genetics* 57S, 755-761

316. Xiao W, Oefner PJ (2001), Denaturing high-performance liquid chromatography: A review, *Hum Mutat.* 17: 439-474

317. Brouwer JR, Willemsen R, Oostra BA (2009), Microsatellite repeat instability and neurological disease, *BioEssays* 31: 71-83

318. Fu YH, Kuhl DP, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJ, Holden JJ, Fenwick RG, Jr., Warren ST, . (1991), Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox, *Cell* 67: 1047-1058

319. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH (1991), Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy, *Nature* 352: 77-79

320. Chakraborty A, Mishra AK, Soni A, Regina T, Mohil R, Bhatnagar D, Bhatnagar A, Chintamani C, Sharma PC, Saxena S (2009), Vitamin D receptor gene polymorphism(s) and breast cancer risk in north Indians, *Cancer Detect.Prev.* 32: 386-394

321. Primer3 plus (2009) Primer 3:

http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi

322. The Wellcome Trust Case Control Consortium (2007), Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls, *Nature* 447: 661-678

323. Esparza-Gordillo J, Weidinger S, Folster-Holst R, Bauerfeind A, Ruschendorf F, Patone G, Rohde K, Marenholz I, Schulz F, Kerscher T, Hubner N, Wahn U, Schreiber S, Franke A, Vogler R, Heath S, Baurecht H, Novak N, Rodriguez E, Illig T, Lee-Kirsch MA, Ciechanowicz A, Kurek M, Piskackova T, Macek M, Lee YA, Ruether A (2009), A common variant on chromosome 11q13 is associated with atopic dermatitis, *Nature Genetics* 41: 596-601

324. Andersson L (2009), Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci, *Genetica* 136: 341-349

325. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001), Linkage disequilibrium in the human genome, *Nature* 411: 199-204

326. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander EA, Ponting CP, Galibert F, Smith DR, deJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin CW, Cook A, Cuff J, Daly MJ, DeCaprio D, Gnerre S, Grabherr M, Kellis M, Kleber M, Bardeleben C, Goodstadt L, Heger A, Hitte C, Kim L, Koepfli KP, Parker HG, Pollinger JP, Searle SMJ, Sutter NB, Thomas R, Webber C, Lander ES (2005), Genome sequence, comparative analysis and haplotype structure of the domestic dog, *Nature* 438: 803-819

327. Sargan DR, Withers D, Pettitt L, Squire M, Gould DJ, Mellersh CS (2007), Mapping the Mutation Causing Lens Luxation in Several Terrier Breeds, *Journal of Heredity* 98: 534-538

328. Mateescu RG, Burton-Wurster NI, Tsai K, Phavaphutanon J, Zhang Z, Murphy KE, Lust G, Todhunter RJ (2008), Identification of quantitative trait loci for osteoarthritis of hip joints in dogs, *American Journal of Veterinary Research* 69: 1294-1300

329. Werner P, Raducha MG, Prociuk U, Sleeper MM, Van Winkle TJ, Henthorn PS (2008), A novel locus for dilated cardiomyopathy maps to canine chromosome 8, *Genomics* 91: 517-521

330. Wiik AC, Wade C, Biagi T, Ropstad EO, Bjerkas E, Lindblad-Toh K, Lingaas F (2008), A deletion in nephronophthisis 4 (NPHP4) is associated with recessive cone-rod dystrophy in standard wire-haired dachshund, *Genome Research* 18: 1415-1421

331. Leegwater PA, van Hagen MA, van Oost BA (2007), Localization of White Spotting Locus in Boxer Dogs on CFA20 by Genome-Wide Linkage Analysis with 1500 SNPs, *Journal of Heredity* 98: 549-552

332. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Anderson N, Biagi TM, Patterson N, Pielberg GR, Kulbokas EJ, Comstock KE, Keller ET, Mesirov JP, von Euler H, Kampe O, Hedhammar A, Lander ES, Andersson G, Andersson L, Lindblad-Toh K (2007), Efficient mapping of mendelian traits in dogs through genome-wide association, *Nature Genetics* 39: 1321-1328

333. Hillbertz S, Nicolette H.C., Isaksson M, Karlsson EK, Hellmen E, Pielberg GR, Savolainen P, Wade CM, von Euler 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, *Nature Genetics* 39: 1318-1320

334. Tian C, Gregersen PK, Seldin MF (2008), Accounting for ancestry: population substructure and genome-wide association studies, *Human Molecular Genetics* 17: R143-R150

335. Devlin B, Roeder K (1999), Genomic control for association studies, *Biometrics* 55: 997-1004

336. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000), Association Mapping in Structured Populations, *The American Journal of Human Genetics* 67: 170-181

337. Wood SH, Clements DN, Ollier WE, Nuttall T, McEwan NA, Carter SD (2009), Gene expression in canine atopic dermatitis and correlation with clinical severity scores, *Journal of Dermatological Science* 55: 27-33

338. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007), PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses, *The American Journal of Human Genetics* 81: 559-575

339. Gentleman R & Ihaka R (2009) R statistical package: http://cran.r-project.org.

340. Teo Y. (2008) Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. *Current Opinion in Lipidology* 19(2), 133-143.

341. Yu K, Wang Z, Li Q, Wacholder S, Hunter DJ, Hoover RN, Chanock S, Thomas G (2008), Population Substructure and Control Selection in Genome-Wide Association Studies, *PLoS ONE* 3: e2551

342. Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, Altshuler D, Ardlie KG, Hirschhorn JN (2005), Demonstrating stratification in a European American population, *Nature Genetics* 37: 868-872

343. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006), Principal components analysis corrects for stratification in genome-wide association studies, *Nature Genetics* 38: 904-909

344. Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, Selmi C, Klareskog L, Pulver AE, Qi L, Gregersen PK, Seldin MF (2008), Analysis and Application of European Genetic Substructure Using 300 K SNP Information, *PLoS Genetics* 4: e4

345. Calboli FCF, Sampson J, Fretwell N, Balding DJ (2008), Population Structure and Inbreeding From Pedigree Analysis of Purebred Dogs, *Genetics* 179: 593-601

346. Laird NM, Lange C (2006), Family-based designs in the age of large-scale gene-association studies, *Nature Reviews Genetics* 7: 385-394

347. Chen WM, Abecasis GR (2007), Family-based association tests for genomewide association scans, *American Journal of Human Genetics* 81: 913-926

348. Cheng H, Ma Y, Ni X, Jiang M, Luo Y, Ying K, Xie Y, Ma Y (2002), Cloning, mapping, and characterization of the human Rab3C gene, *Biochemical Genetics* 40: 263-272

349. Hoffjan S, Stemmler S (2007), On the role of the epidermal differentiation complex in ichthyosis vulgaris, atopic dermatitis and psoriasis, *British Journal of Dermatology* 157: 441-449

350. Piekutowska A, Pin D, RΦme CA, Gatto H, Haftek M (2008), Effects of a Topically Applied Preparation of Epidermal Lipids on the Stratum Corneum Barrier of Atopic Dogs, *Journal of Comparative Pathology* 138: 197-203

351. Myers JC, Li D, Amenta PS, Clark CC, Nagaswami C, Weisel JW (2003), Type XIX collagen purified from human umbilical cord is characterized by multiple sharp kinks delineating collagenous subdomains and by intermolecular aggregates via globular, disulfide-linked, and heparin-binding amino termini, *Journal of Bioogical.Chemistry* 278: 32047-32057

352. Carr CS, Sharp PA (1990), A helix-loop-helix protein related to the immunoglobulin E box-binding proteins, *Molecular Cellular Biology* 10: 4384-4388

353. Gupta S, Barrett T, Whitmarsh AJ, Cavanagh J, Sluss HK, Derijard B, Davis RJ (1996), Selective interaction of JNK protein kinase isoforms with transcription factors, The *EMBO Journal* 15: 2760-2770

354. Uniprot (2009) MAPK8 http://www.uniprot.org/uniprot/P45983#ref2. Accessed Sep 27, 2009

355. Uniprot (2009) SUCLG2 involved in carbohydrate metabolism. http://www.uniprot.org/uniprot/Q96I99. Accessed Aug 3, 2009

356. Uniprot (2009) FAM19A4 http://www.uniprot.org/uniprot/Q96LR4.. Accessed Sep 27, 2009

357. Rajaiya J, Nixon JC, Ayers N, Desgranges ZP, Roy AL, Webb CF (2006), Induction of immunoglobulin heavy-chain transcription through the transcription factor Bright requires TFII-I, *Molecular Cellular Biology* 26: 4758-4768

358. Uniprot (2009) FSCN1. http://www.uniprot.org/uniprot/Q16658. Accessed Sep 27, 2009

359. Howell MD, Kim BE, Gao P, Grant AV, Boguniewicz M, DeBenedetto A, Schneider L, Beck LA, Barnes KC, Leung DYM (2007), Cytokine modulation of atopic dermatitis filaggrin skin expression, *Journal of Allergy and Clinical Immunology* 120: 150-155

360. McGrath JA (2008), Filaggrin and the great epidermal barrier grief, *Australasian Journal of Dermatology* 49: 67-74

361. Uniprot (2009) SLC35F3 http://www.uniprot.org/uniprot/Q8IY50. Accessed Sep 27, 2009

362. Dale BA, Scofield JAH, Hennings H, Stanley JR, Yuspa SH (1983), Identification of Filaggrin in Cultured Mouse Keratinocytes and Its Regulation by Calcium, *Journal of Investigative Dermatology* 81: 90s-95s

363. Gur G, Rubin C, Katz M, Amit I, Citri A, Nilsson J, Amariglio N, Henriksson R, Rechavi G, Hedman H, Wides R, Yarden Y (2004), LRIG1 restricts growth factor signaling by enhancing receptor ubiquitylation and degradation, *The EMBO Journal* 23: 3270-3281

364. Kaufman L, Hayashi K, Ross MJ, Ross MD, Klotman PE (2004), Sidekick-1 is upregulated in glomeruli in HIV-associated nephropathy *Journal of the American Society of Nephrology* 15: 1721-1730

365. Hon KLE, Lam MCA, Wong KY, Leung TF, Ng PC (2007), Pathophysiology of nocturnal scratching in childhood atopic dermatitis: the role of brain-derived neurotrophic factor and substance P, *British Journal of Dermatology* 157: 922-925

366. McCarthy MI, Hirschhorn JN (2008), Genome-wide association studies: potential next steps on a genetic journey, *Human Molecular Genetics* 17: R156-R165

367. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, Hirota Y, Mori H, Jonsson A, Sato Y, Yamagata K, Hinokio Y, Wang HY, Tanahashi T, Nakamura N, Oka Y, Iwasaki N, Iwamoto Y, Yamada Y, Seino Y, Maegawa H, Kashiwagi A, Takeda J, Maeda E, Shin HD, Cho YM, Park KS, Lee HK, Ng MCY, Ma RCW, So WY, Chan JCN, Lyssenko V, Tuomi T, Nilsson P, Groop L, Kamatani N, Sekine A, Nakamura Y, Yamamoto K, Yoshida T, Tokunaga K, Itakura M, Makino H, Nanjo K, Kadowaki T, Kasuga M (2008), Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus, *Nature Genetics* 40: 1092-1097

368. Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, Andersen G, Ng DPK, Holmkvist J, Borch-Johnsen K, Jorgensen T, Sandbaek A, Lauritzen T, Hansen T, Nurbaya S, Tsunoda T, Kubo M, Babazono T, Hirose H, Hayashi M, Iwamoto Y, Kashiwagi A, Kaku K, Kawamori R, Tai ES, Pedersen O, Kamatani N, Kadowaki T, Kikkawa R, Nakamura Y, Maeda S (2008), SNPs in KCNQ1 are associated with

susceptibility to type 2 diabetes in East Asian and European populations, *Nature Genetics* 40: 1098-1102

369. Nalls MA, Wilson JG, Patterson NJ, Tandon A, Zmuda JM, Huntsman S, Garcia M, Hu D, Li R, Beamer BA, Patel KV, Akylbekova EL, Files JC, Hardy CL, Buxbaum SG, Taylor HA, Reich D, Harris TB, Ziv E (2008), Admixture Mapping of White Cell Count: Genetic Locus Responsible for Lower White Blood Cell Count in the Health ABC and Jackson Heart Studies, *The American Journal of Human Genetics* 82: 81-87

370. Reche PA, Soumelis V, Gorman DM, Clifford T, Liu M, Travis M, Zurawski SM, Johnston J, Liu YJ, Spits H, de Waal MR, Kastelein RA, Bazan JF (2001), Human thymic stromal lymphopoietin preferentially stimulates myeloid cells, *Journal of Immunology* 167: 336-343

371. Zhang W, Wang J, Wang Q, Chen G, Zhang J, Chen T, Wan T, Zhang Y, Cao X (2001), Identification of a novel type I cytokine receptor CRL2 preferentially expressed by human dendritic cells and activated monocytes, *Biochemical and Biophysical Research Communications* 281: 878-883

372. Gillingham AK, Pfeifer AC, Munro S (2002), CASP, the alternatively spliced product of the gene encoding the CCAAT-displacement protein transcription factor, is a Golgi membrane protein related to giantin, *Molecular Biology of the Cell* 13: 3761-3774

373. Batty IH, van der KJ, Gray A, Telfer JF, Dixon MJ, Downes CP (2007), The control of phosphatidylinositol 3,4-bisphosphate concentrations by activation of the Src homology 2 domain containing inositol polyphosphate 5-phosphatase 2, SHIP2, *The Biochemical Journal* 407: 255-266

374. Gudjonsson JE (2008), Genetic variation and psoriasis, *Giornale.Italiano di Dermatologia e Venereologia:organo ufficiale societa italiana di dermatologia e sifilografia* 143: 299-305

375. Nam D, Kim SY (2008), Gene-set approach for expression pattern analysis, *Briefings in Bioinformatics* 9: 189-197

376. Borgatti SP, Mehra A, Brass DJ, Labianca G (2009), Network Analysis in the Social Sciences, *Science* 323: 892-895

377. Almaas E (2007), Biological impacts and context of network theory, *Journal of Experimental Biology* 210: 1548-1558

378. Holden M, Deng S, Wojnowski L, Kulle B (2008), GSEA-SNP: Applying Gene Set Enrichment Analysis to SNP data from genome-wide association studies, *Bioinformatics* 24(23):2784-5

379. Joo J, Kwak M, Ahn K, Zheng G (2009), A Robust Genome-Wide Scan Statistic of the Wellcome Trust Case-Control Consortium, *Biometrics*

380. Torkamani A, Topol EJ, Schork NJ (2008), Pathway analysis of seven common diseases assessed by genome-wide association, *Genomics* 92: 265-272

381. Browning BL, Browning SR (2008), Haplotypic analysis of Wellcome Trust Case Control Consortium data, *Human Genetics* 123: 273-280

382. Elbers CC, van Eijk KR, Franke L, Mulder F, van der Schouw YT, Wijmenga C, Onland-Moret NC (2009), Using genome-wide pathway analysis to unravel the etiology of complex diseases, *Genetic Epidemiology* 33: 419-431

383. Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, Carter NP, Hurles ME, Feuk L (2007), Challenges and standards in integrating surveys of structural variation, *Nature Genetics* 39(7 suppl):s7-15

384. Henrichsen CN, Chaignat E, Reymond A (2009), Copy number variants, diseases and gene expression, *Human Molecular Genetics* 18: R1-R8

385. Karlsson EK, Lindblad-Toh K (2008), Leader of the pack: gene mapping in dogs and other model organisms, *Nature Reviews Genetics* 9: 713-725

386. Presland RB, Haydock PV, Fleckman P, Nirunsuksiri W, Dale BA (1992), Characterization of the human epidermal profilaggrin gene. Genomic organization and identification of an S-100-like calcium binding domain at the amino terminus, *Journal of Biological Chemistry* 267: 23772-23781

387. The Malaria Genomic Epidemiology Network (2008), A global network for investigating the genomic epidemiology of malaria, *Nature* 456: 732-737

388. Verra F, Avellino P, Bancone G, Mangano V, Modiano D (2008), Genetic epidemiology of susceptibility to malaria: not only academic exercises, *Parassitologia* 50: 147-150

389. Jallow M, Teo YY, Small KS, Rockett KA, Deloukas P, Clark TG, Kivinen K, Bojang KA, Conway DJ, Pinder M, Sirugo G, Sisay-Joof F, Usen S, Auburn S, Bumpstead SJ, Campino S, Coffey A, Dunham A, Fry AE, Green A, Gwilliam R, Hunt SE, Inouye M, Jeffreys AE, Mendy A, Palotie A, Potter S, Ragoussis J, Rogers J, Rowlands K, Somaskantharajah E, Whittaker P, Widden C, Donnelly P, Howie B, Marchini J, Morris A, SanJoaquin M, Achidi EA, Agbenyega T, Allen A, Amodu O, Corran P, Djimde A, Dolo A, Doumbo OK, Drakeley C, Dunstan S, Evans J, Farrar J, Fernando D, Hien TT, Horstmann RD, Ibrahim M, Karunaweera N, Kokwaro G, Koram KA, Lemnge M, Makani J, Marsh K, Michon P, Modiano D, Molyneux ME, Mueller I, Parker M, Peshu N, Plowe CV, Puijalon O, Reeder J, Reyburn H, Riley EM, Sakuntabhai A, Singhasivanon P, Sirima S, Tall A, Taylor TE, Thera M, Troye-Blomberg M, Williams TN, Wilson M, Kwiatkowski DP (2009), Genome-wide and fine-resolution association analysis of malaria in West Africa, *Nature Genetics* 41: 657-665

390. Schoenborn JR, Dorschner MO, Sekimata M, Santer DM, Shnyreva M, Fitzpatrick DR, Stamatoyonnapoulos JA, Wilson CB (2007), Comprehensive epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferon-[gamma], *Nature Immunology* 8: 732-742

Appendix 1 RNA concentration and RIN values obtained by Bioanalyser

Sample Name	Type	Storage	RNA Concentration	RNA Integrity Number (RIN):
206080L	Lesional	snap frozen	524.4	9.3
206462nl	Non-lesional	snap frozen	213.8	5.8
206556L	Lesional	snap frozen	536.7	4.9
206821L	Lesional	snap frozen	468.6	5.5
206821nl	Non-lesional	snap frozen	216.8	6.2
207616L	Lesional	snap frozen	248.9	5.2
207616NL	Non-lesional	snap frozen	63.3	7.2
207632L	Lesional	snap frozen	80.7	4.9
207632NL	Non-lesional	snap frozen	105.7	6.6
207676L	Lesional	snap frozen	205.2	5.1
207676NL	Non-lesional	snap frozen	64.3	5.9
207935L	Lesional	snap frozen	1020.1	5.0
207935nl	Non-lesional	snap frozen	446.4	4.6
208462L	Lesional	snap frozen	503.4	6.0
208511L	Lesional	snap frozen	329.9	5.7
208511nl	Non-lesional	snap frozen	140.4	6.4
208628L	Lesional	snap frozen	436.5	5.9
208628nl	Non-lesional	snap frozen	99.7	6.3
208781L	Lesional	snap frozen	798.3	5.4
208781nl	Non-lesional	snap frozen	303.8	5.9
208829L	Lesional	snap frozen	729.6	4.8
208829nl	Non-lesional	snap frozen	228.3	5.2
A10L	Lesional	snap frozen	259.5	8.3
A10NL	Non-lesional	snap frozen	29.0	7.9
A11L	Lesional	snap frozen	265.3	8.2
A12L	Lesional	snap frozen	25.8	9.4
A12NL	Non-lesional	snap frozen	25.8	9.5
A13L	Lesional	snap frozen	240.0	7.7
A13NL	Non-lesional	snap frozen	119.0	8.1
A14NL	Non-lesional	snap frozen	119.1	8.4
A16L	Lesional	snap frozen	168.1	8.0
A16NL	Non-lesional	snap frozen	79.6	9.0
A17L	Lesional	snap frozen	224.1	8.3
A18L	Lesional	snap frozen	119.1	7.3
A18NL	Non-lesional	snap frozen	37.4	7.3
A1L	Lesional	snap frozen	45.6	8.9
A1NL	Non-lesional	snap frozen	2252.7	8.2
A2L	Lesional	snap frozen	85.5	6.1
A2NL	Non-lesional	snap frozen	91.5	8.0
A3L	Lesional	snap frozen	1614.6	7.2
A3NL	Non-lesional	snap frozen	87.5	7.9

Appendix 1 continued

A4L	Lesional	snap frozen	73.2	8.3
A4NL	Lesional	snap frozen	42.1	8.4
A6L	Lesional	snap frozen	36.8	8.1
A6NL	Non-lesional	snap frozen	94.3	7.9
A7NL	Non-lesional	snap frozen	74.3	8.1
A8L	Lesional	snap frozen	192.7	8.2
N10	Control	snap frozen	58.1	2.4
N11	Control	snap frozen	107.2	7.5
N13	Control	snap frozen	55.2	2.4
N14	Control	snap frozen	35.4	2.5
N15	Control	snap frozen	18.6	9.0
N3	Control	snap frozen	25.4	6.1
N4	Control	snap frozen	254.4	7.6
N5	Control	snap frozen	83.0	7.5
unknL	Lesional	snap frozen	353.4	6.4
unknNL	Non-lesional	snap frozen	152.9	6.4
206140nl	Non-lesional	rnalater	267.5	9.0
210817nl	Non-lesional	rnalater	266.9	8.7
210488nl	Non-lesional	rnalater	1088.1	8.5
cbNL	Non-lesional	rnalater	21.5	8.5
206140L	Lesional	rnalater	46.0	8.4
210817L	Lesional	rnalater	452.5	7.3
210488L	Lesional	rnalater	69.1	7.3
210534L	Lesional	rnalater	708.4	6.6
210479L	Lesional	rnalater	54.1	6.1
Control 2	Control	rnalater	59.2	9.0
Control 4	Control	rnalater	649.3	8.9
MARVIN	Control	rnalater	653.6	8.7
Control 6	Control	rnalater	1234.7	8.5
Control 8	Control	rnalater	1332.7	8.4
Control 5	Control	rnalater	1105.0	8.3
Control 3	Control	rnalater	707.2	8.3
CLYDE	Control	rnalater	1224.5	8.2
Control 7	Control	rnalater	1155.5	8.1
edin 1	Control	rnalater	1846.1	8.0
Control 1	Control	rnalater	87.0	8.0
HELEN	Control	rnalater	2136.3	7.7
Control 9	Control	rnalater	949.6	7.6

Appendix 2 Slope, R^2, percentage efficiency and E values for each qPCR assay

Gene	Slope	R^2	% efficiency	E value
SCCA2	-3.477	0.99	94	1.88
Cystatin A	-3.316	0.99	100	2
Card 4	-3.445	0.99	95	1.9
P-selectin	-3.441	0.99	95	1.91
PKP2	-3.415	0.99	96	1.93
PPARɣ	-3.195	0.98	106	2.11
SGPL1	-3.247	0.99	103	2.06
TNF α	-3.268	0.98	102	2.05
Cadherin 13	-3.211	0.99	105	2.10
CMA1	-3.440	0.99	95	1.91
DPP4	-3.505	0.99	93	1.86
POSTN	-3.505	0.99	93	1.86
SAA-1	-3.469	0.99	94	1.88
SPINK5	-3.366	0.99	98	1.96
ARTS-1	-3.329	0.99	100	1.99
INPPL1	-3.388	0.98	97	1.95
S100A8	-3.471	0.99	94	1.88
Cullin4A	-3.494	0.99	93	1.87
STAT2	-3.434	0.99	96	1.91
TIMP1	-3.317	0.99	100	2

Appendix 3 WAVE methods: Fragments and Temperatures

CMA1 Promoter region

CAGATTCCCCCAGTCACAAGCTCAAGCTAAGACCAAGGGCATGGTAGGATGATTTAGGGAAGGTA
TGTTTTCCCTGTGATTCCTTATCTCTGAATTAAATGGCTTCTGAATCTCCAAAAGAGTGGCGCCTGG
GTGGCTCAGTCAGTTAAGCATCTGCCTTTGACTCAGGTCATGATCCTGGGGTTCTGGGAGAAAGCT
CCATGTCAGGCTCCTTGTCAGCATGGAATCTGCTTCTCCCTCTTCCTCTGCAGCTCCTCGTGATTGT
TTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGTCAAATAAATAAATAAAATATAA
TAAATCTTCAAGAGAAACGACACTCTGCCTGTT

Temps: 50, 57.5, 60.5
Fwd: CAGATTCCCCCAGTCACAAG
Rev: TAACAGGCAGAGTGTCGTTTC
Product size: 375bp

CMA1 5th exon

TTGAAGGAGAAGGCCAACCTGACCCTGGCCGTGGGGACCCTCCCCCTCTCACCCCAGTTCAACTTC
GTCCCACCTGGGAGAATGTGCCGGGTGGCTGGCTGGGGAAAAAGACAAGTCAATGGATCAGGCTC
TGACACTCTGCAAGAGGTGAAGCTGAGGCTCATGGACCCCCAGGCCTGTAGACACTACATGGCTTT
TGACCA

Temps: 62.3, 64.2
Fwd: TTGAAGGAGAAGGCCAACC
Rev: TGGTCAAAAGCCATGTAGTGTC
Product size: 203bp

CMA1 6th exon

CTCTTCTGTGTGCTGGGGTAGCCCAGGGAATTGTGTCCTATGGGCAGAATGATGCAAAGCCCCCTG
CTGTCTTCACCCGAATCTCCCACTACCGGCCCTGGATCAATAAGGTTCTGAAGCAGAATAAAGCCT
GAATGGAGCCTGGGCAAGCCTGAGGGGAAATCTGGAACCAGACCTGAGCAGGCTCTCTGCCACTT
ACTCTGGAGCTGCCTCTAGTCTCTACTGAGGCCCCACTACATCCCTCAGACCC

Temps: 61.4
Fwd: CTCTTCTGTGTGCTGGGGTAG
Rev: GGGTCTGAGGGATGTAGTGG
Product size: 250bp

SGPL1 2nd exon

CATTGAAGAAGTGAGTGAAGAACAGACTACCATCAGATGAACTTGTTTTGTACTGGACACTGTGG
GGGCTTGAAATGATGACTAAGGTCTGATCAGTGTTCAAGGACTCAACCACGTAGTGCGGAAGATA
AAAACAGATGAGTAGGCTGACTGGTGATAGGCTCTGTGTGGTGGCACAAGGAGATACTTCAGGCG
AGTTTAATGGG

Temps: 59.8
Fwd: CATTGAAGAAGTGAGTGAAGAACAG

Rev: CCCATTAAACTCGCCTGAAG
Product size: 206bp

SGPL1 6th exon

GTGCTCTCACCTGTGTGCTTCTCCCCCTCAGTATGGCTACGCCCCCAAAGGCTCTTCCGTGCTGTTA
TACAGTGACAAGAAGTACAGGAGCCATCAGTTCTTCGTGGCGACTGACTGGCAGGGCGGCATCTA
CGCTTCCCCGACCATCGCGGGCTCACGGCCCGGAGGCATTAGCGCGGCCTGTTGGGCTGCCTTGAT
GCACTTCGGTGAGAGCGGCTATGTTGAAGCCACCAAACAGATCATCAAAACTACTCGCTTCCTCAA
GTCAGAGTATGTGTGGAGGACTGGTTCGGC

Temps: 59.2, 61.2, 65.7
Fwd: GTGCTCTCACCTGTGTGCTTC
Rev: GCCGAACCAGTCCTCCAC
Product size: 294bp

SGPL1 14th exon

ACCGCCAGTAACGACCTCTCACTGGATTTGTAGGGTGCGATCTATGGCATGGCCCAGACGACAGTT
GACAGGAACCTGGTCGCAGAATTGTCCTCGGTCTTCTTGGACAGCCTCTTCAGCACGGACACTGTG
ACTCCAAGCAGCCAGATGAATGGTTCTCCGAAACCCCGCTGAGCCTGGATACTCTGTGACCCTGA

Temps: 62.4
Fwd: ACCGCCAGTAACGACCTC
Rev: TCAGGGTCACAGAGTATCCAG
Product size: 197bp

DPP4 Promoter region (p1)

ACAAACAAACAGCAACCCAAGGCCCTTTCTTCAATGAATACGTGGCCTAGCTGAGTACACCACAC
TGGTAGGCATGGAAAGATAGTCATTAACACAAGTTATTAATTACAAAAGCCAAATGAGTAGTGTG
AGGGCAGAAGTCAAAGGAGGCAGTATTGAAGTCTCCATGAGAGAGGAGGGAATTGGCAGGTGGA
GGACTGTGACAGCCCGGGAAGCAGTTGGAGATGCTGCAGGGAAGAGGCTGAGGAAGAGAGGCAA
AGACAGAAAAGACAAAATTGGATGCAGGTGGTAGTAAAAGAGAGGGAGACATCAAAAATGATTT
TCAGACTTTGAATTTGGATCTCCAGTAACAGAATCTAAAACCACACAAATTCCATAGGGAACATTA
TTTTGGAAGGAATTATTTTAGACAAGGTAAGCCTAGGGTTAGCATGGGACAAATCAACTAAAAAG
GCCCAACACAGCACCAGAAATGTGAGTGTGGGGTGTGGAA

Temps: 56,61
Fwd: ACAAACAAACAGCAACCCAAG
Rev: TTCCACACCCCACACTCAC
Product size: 493bp

DPP4 promoter region (p2)

GTGTGGAAGGCTGAGGTAATGAGTATAGGTTTTGTCATGTATTCAAATGGAAAATATGCCTGAAGC
CCAGGAAGTCATATTCTCTAGAGTCCTCTAGAAGGAGAAATGGAGGCAGAGAAGGGCAAAGAAG
CTGATGCTCTGGGCATTTCTACCTTAGAGGTTAATAGAGAAAAATAAAGTCAAAGGAGGATGGAA
AATTAAGGGTCAGAGGTCAGGAGAAGTGAGGTGCCCTAGAATCCAAAGGAGAGTTTCATAAATCA
CCACCCTCCAAGTTCAAG

Temps 57.5
Fwd: GTGTGGAAGGCTGAGGTAATG
Rev: CTTGAACTTGGAGGGTGGTG
Product size: 278bp

DPP4 6th exon

CAAGCACTCCTCATTTGATGCAGAAAAACCACAAAGGCTAACAATTTGGGATTTTTCACAGATATG
CAGGCCCATGTAGTCAAAAAGCAGACGCTGTCTTCAGACTCAACTGGGCTACTTACCTTGCAAGCA
CAGAAAACATTATCGTAGCTAGCTTTGATGGCAGAGGAAGTGGTTACCAAGGAGATAAGATCATG
CACGCCGTCAACAGAAGACTGGGAACATTTGAAGTTCAAGATCAAATTGACGCAGCCAGGTGAGT
GACTAGGAGATGAATCCGAAGGTCA

Temp: 59.5
Fwd: CAAGCACTCCTCATTTGATGC
Rev: TGACCTTCGGATTCATCTCC
Product size: 287bp

DPP4 21st exon

CAAGCACTCCTCATTTGATGCAGAAAAACCACAAAGGCTAACAATTTGGGATTTTTCACAGATATG
CAGGCCCATGTAGTCAAAAAGCAGACGCTGTCTTCAGACTCAACTGGGCTACTTACCTTGCAAGCA
CAGAAAACATTATCGTAGCTAGCTTTGATGGCAGAGGAAGTGGTTACCAAGGAGATAAGATCATG
CACGCCGTCAACAGAAGACTGGGAACATTTGAAGTTCAAGATCAAATTGACGCAGCCAGGTGAGT
GACTAGGAGATGAATCCGAAGGTCA

Temps: 59.5
Fwd: CAAGCACTCCTCATTTGATGC
Rev: TGACCTTCGGATTCATCTCC
Product size: 287bp

SPRR1B Promoter region (p1)

CACTTCTCCCCCAACACAAGCCCAAAGCCTGGGTCTAAGGCAGAGGGTGGCACTGGGGATTAGGC
ATCAGTGACATCATTTGTCACACAGCCAGTGCCAAAAAATATCAAGTGGGTGGTTTTCATCACAAA
AGCTGATCCAACTGGTAATTGAGGGGAATAGGGTGAATCATGTTTGACAAGTAAGGAATGAAAAC
AGAGGGGGCCTGGGGTGATTCTGTTTCCTTGAGGCAGGGCTCATGCATCCTATAAAAAGCTAGCTG
GCCACTGACTTTACATCAGATCAGAGCAACCAGGCACCCCTTTCTCTGTGTACCAGGTAAGTCTCT
GATTGGAATCAATTTGAAATTCATGTTTATTTTTTTTATTATTGGAAACTGGACAAAGGCAAACCTG
TGTAAGAGGGCTGGTGAGG

Temps: 58.5
Fwd: CACTTCTCCCCCAACACAAG
Rev: CCTCACCAGCCCTCTTACAC
Product size: 414bp

SPRR1B Promoter region (p2)

CCCTTTCAGCAACACCACTCTTTTCTCTGTGTATGACACCTATAAAAATGTTGATTCATGAGATGGA
CCTCTGTTGTGCATTTGTTATATCTCAGGCACCATGCTAGTCACTTTGGGTAGATCAGACCATTATG
GAACTTACTCAGCAAGCTCACAGCCTAGCAGGGAAGATAATGCATGCTTATACATAGGTGAAATA
CAAAGCAGATACTGACTGGTATGGGAACAGAGGGACAGAGAAGGGCCATGGCAGTTTGGGGGAG
CTATGCTCAGGATCTCCTGGTACAAGATATGAGCTATAATGGAGGCATATCTGGACAAGCAAAAT
GGGAAGGAGAGCATTTCAGAGCAGATTGCATAAGCAAAGGCAGAATGAGGTCACAGGAGCTGAG
ACTCACAGAGGCATATTTAAGTCACAGGTATGCCACATGGTC

Temps: 57.9
Fwd: CCCTTTCAGCAACACCACTC
Rev: GACCATGTGGCATACCTGTG
Product size: 434bp

SPRR1B Exon 1

AGAAGCAACCCTGCATCCCTCCCCCCCAGCCTCAGCAGCAGGTGAAACAGCCATGCCAGCCTCCA
CCCCAGGAGCCATGTGCCCCCAAAACCAAGGAGCCATGCCACCCCAAGGTTCCAGAGCCTTGCCA
CACCAAGATTCCAGAGCCCTGCCACACCAAGGTTCCAGAGCCCTGCCAGCCCAAGGTTCCAGAGC
CATGTCCCTCAACAGTGAC

Temps: 63.8
Fwd: AGAAGCAACCCTGCATCC
Rev: GTCACTGTTGAGGGACATGG
Product size: 214bp

SPRR1B 3rd exon

CAACAGGTGAAACAGACTTGCCAGCCTCTACCTCAGGAGCCATTTGTTCCCATAACCAAGAATCCA
TGTGACACCAAGTTACCGCAACCTGGGAACACCAAACTTCCTGAACCAAACTACCCTAAATATCCT
GAGCCAGGCCATCCCATAATTCCTGAATCAGGCCACATCAAGGTCCCCGAGCCAGCCTACTCTGAG
GTCCTTCAGCCTGGCTACACCAAGGTACCAGAACCAAGTCCCTCAACCATTATTCCAGGCCAGAAG
ACCAAGCAGAAATGATGTGGTCCTTAGCCATGACCT

Temps: 59.8, 61.8
Fwd: CAACAGGTGAAACAGACTTGC
Rev: AGGTCATGGCTAAGGACCAC
Product size: 300bp

S100A8 1st & 2nd exons

CCCATGTGTTTCGTGTTGCAGGTGGGGCAGATCCTTGGGCACCATGCTGACGGAACTGGAGAGTGC
CATAAACTCCCTCATTGAGGTTTACCACAAGTACTCCCTGGTGAAGGGGAATTACCACGCCCTCTA
CAGAGATGACTTGAAGAAATTGTTAGAGACAGAGTGTCCTCAGTACATGAAGGTGAGGAGGGGCT
GGGTGTGGCAGGGGTCTCTGTCTGGCCAATGAAGCCCTGGGTCCCTGGTCCCCTACGCCACCCTAT
GCAGACAGCCATGGCTCAGCTCTTGCCAAGGTCTTTACGCACTGGCTTCTTTCTCCTCCATCTTGAC
AGAAAAAGGATGCAGACACCTGGTTCCAAGAGTTGGATGTCAATAGCGATGGTGCAATTAACTTC
GAGGAGTTCCTCATATTGGTGATAAAGGTGGGCGTGGCCTCCCACAAAGACATCCACAAGGAGTA
GCAGAGCCCTGGGCCCCGGGGCTGTCCCTGCCGAATAATA

Temps: 59.5, 61
Fwd: CCCATGTGTTTCGTGTTGC
Rev: TATTATTCGGCAGGGACAGC
Product size: 500bp

S100A8 Promoter region (p2)

GGGATTAGCAGCAGCAAGAGAGGAGGACTTGAGAGTGGGCATCCAGGCTGGCTGAGCCAAATGA
TGCCAGTGGCCCAGCAAAGGAGTCCTCACCATGGCCTCTTCCTGATTTATCCTGCTGGGCACTACT
GCCTAGCAAATACCTTCCTCTTTCCATTTTTCCTACTTCCCCACCTCCCATCAAAATTTTTCATTCCG
TGCAGTGATTGCCCCATTCACCTGGTTGAGAAACCAGAGGCTGTGGCAACCTGGGGAGGGAGGGG
CTATCTCTGCATGGCCTAGGGCCCTGGCAGCTGGCCACAGCCTGGCTGCTATAAAGGGGAGCTGCC
ACTGAGCCCTGCATATCCTCTGTCAACTCTGTTTCGGGAAACCTGGTAAGTAAGACTGCCTCAGTC
TGCTCCATGCTTTGGGGCTGCCCCCCCACCCCGGGAGGGCCAGAGGTGGAGCAGAGAGGGGAGAG
AAAGCTAAGGGAGGTCTGGA

Temps: 60, 62
Fwd: GGGATTAGCAGCAGCAAGAG
Rev: TCCAGACCTCCCTTAGCTTTC
Product size: 480bp

IL1RL1 10th exon

CACAGCAACAAGAAATCTGACCCGTGTCCTTTATTCCTAGATGCAGCCACCGCAGTGGAAACCAA
CATACGAAAGAGTAGGCGCCACATTTTTATCCTGACTCCTGAGATCACACACAGCAGGGAGTTTGC
CTACGAGCAGGAGATCGCCCTGCACAGTGCCCTCATCCAGAACGACGCCAAGGTGATTCTTATAG
AATTGGAGGCTCTGAGCAAGCCAGGTGGACTACAGTTTGGGGAGCTTCAAGATTCCCTCAAGCAC
CTTGTGGAAGTGCAAGGTACCATCAAGTGGAGGGAAGACCATGTGGCCAACAAACAGTCGCTGAA
TTCCAAATTCTGGAAGCATGTGAGGTACCACATGCCTGTGCCCAACAGACTGCCAGGGAAGACAT
CCAGTTTGGCTTCCCTGAGTGCCCAGGGGCAGTAGTGCTTGCTTTGATGTGCTAATGCACCAGATT
TCCG

Temps: 60.4
Fwd: CACAGCAACAAGAAATCTGACC
Rev: CGGAAATCTGGTGCATTAGC
Product size: 461bp

IL1RL1 Promoter region (p1)

TGCCAAATGAAGTGTTGAGGGATATCTGGAAAGGCCAACCCAAAGCGCCAGGATTTCAGTATCAT
TGAGAACCCGATGAGTAATAATATTAGGACAATAAACAGCCAGTGTTATTTAGTTTGAGAAACTG
ATGCTAAAATAAGAGGAGGTAGTTTAATTATTTCCTCTTAAGGCATTACTCAGTGTTGTCCTTATGT
TTAAATATTTATGTACCAAGGTCAAGAATTCTTGGTACATGATGCACCAGGATTTTTGAACAGTCA
TAAATCGTGGCTAAAAATCTAACTTCAGTGTGGGTGGAGTGATGAATTTCAGGACGTCTTTGAGTG
TCTCCTCTTCAAAATATAAGAGAGAGGCCATACACCTGAGGAGTCTTCAGGGCAAGAGACTCGGG
AGCTGGAAACTATTCCTTGCTCAGTCACAGACTTGGATTTTATTCCCTTCTGCCTTTCAGTTTGG

Temps: 55, 56.8 and 59
Fwd: TGCCAAATGAAGTGTTGAGG
Rev: CCAAACTGAAAGGCAGAAGG
Product size: 459bp

IL1RL1 Promoter region (p2)

TTGGTGAGTCTTGGGGACACGCTGAGGCAGGGACACATTTTGTGTAGTGTCTAGGGCAATCACTTG
ATTAACACTTTACCAATCACTTTCCTATATTTTTAGTTGCTTTGTGATGACTATTTGTATATGTGTCT
GCATGTGGGTCACACATGTGTGCAGGCAAGTCCCCAGCCTATGTGCATATCTGTGGATTTATGTTA
GCATTTATAACATTGTATTTTCTAGGGTCTGTTGAAATATAGTTTTAATTTGTTTCAAAAATGACAA
CTACTTTTCTTTCCTCCTAATTCAGCTTCTTCCAGTCCATTCTCATGATACTATGAAATAGTACAACT
ATTGCAAGGCCTCTGTTGTCCAAAGGTTCAGTAGGAATAAATGCAATGTTTCATATCTGAAACCCT
GCTGTTATACATATTTCTACCTAATGAATACCTGTTGATTGATGAATTCTCCTAGGGGGTATTGGAC
AAGTAAAAATCATGGCAATTTTGCTTGTGATTCTGTGTGATTTATGTTTCAGTGGAGACACTAAAA
CTCTTTCTAACTGTAAATCTCAGGAGGCACCATCAACCT

Temps: 55 and 56.5
Fwd: TTGGTGAGTCTTGGGGACAC
Rev: AGGTTGATGGTGCCTCCTG
Product size: 573bp

MS4A2 Promoter region

AAGATGGCACGGGTTTAATGCATTGCTGTCAGAAAGGTGCACTCATATGTCTTCTGGTGTCCTCTT
GCTGTACGTAGCCTAGCTTCCCGGAGAAGGAGGAGAAGTTTCATCACATCCCAGGGAGGTAATAC
ATGTGCAATTTACAGTGTCTTACAGTGTAGGGAAGAGCCACTCAAGGTTATGTCTAAATTCAACTG
AAACAAAAGGACCCGGAATCTGTGAATCTCACCTGTGGATTCATTTCACCACGTATTCCAATTCTT
GTAAGTTATTCATGCTATTAACATGAACATGTCCGTTAAGCTCTGACCCTATTCAGATGCCCTGCTC
CATCCAGGGCGGCCTAGATCTACTCACAGGATCATCAGGCTCTTGTCCCTTGGTGCTTCGTGGCCA
CAGCTCTGCCATCGTCAAAATCA

Temps: 58 and 61
Fwd: AAGATGGCACGGGTTTAATG
Rev: TGATTTTGACGATGGCAGAG
Product size: 419bp

MS4A2 5th & 6th exon

ACACACAGGGCTATTGAACGCACTGGCAGGTGGGGGCTCAGGCCTGCCTGCTGTCCTTCACCATCA
GGCGTGGGGCAGCCTGGGAGCCAACACTGTCAGCAGCATAGCTGCGGGAATAGGAATCTTTATCC
TGACGGTCAACCTGCAGAGGAGCTCGGCTTACATCCACAACTGCCAGCAAGCTCCCAGAGACGAC
TTCTGCTTTGTGGCTTGTTTTTCCACAGTACGTATTTCTTGTGAGGGGGCTAAGGTCTGGGTCCTAC
TCTAAGTACAAGCCCTCTCCTTTCCCACGAACGTCGTCCTTTGCTGTAACCTCTATCACATAGTGTA
GTGTTCGGGGGGTTCATGGACAAAGATGCTGGCTGCCGCTCCCTCCCCAGCTGGGTTCTGTGAGGC
TGGAGTCACCCGCCTTGACCTCCCTTTTTGTGGCTTTCCCACGTATGTGTCACTTCATTGGGAGTGA
TCACATCGGGGTGATGAAATCAGGGGCTTCTGTTAGGAACACACCTTTGGGGCACCTGCCTTGTGA
CTCTACTTGTCTTGTTGGTTACTGTTCAACAGGAAATTGTGGCTATGATCCTGTTTCTCACTATTCTG
GGGTTTTGCAGTGCTGTGTCACTGACGATCTATGGAGTTGGAGAATTAGTCCAAGGGAACAAGGT
AAGTAGAGGCCTGATAATCCTGAATGACAAGCTGAAGC

Temp: 58, 60, 62
Fwd: ACACACAGGGCTATTGAACG
Rev: GCTTCAGCTTGTCATTCAGG
Product size: 700bp

P-selectin 5th exon

AAGGGCTCATGCTATTGCTCTGTGGCTGTTTGTAGCTACCCAGTGTCCACCCCTAAAGACTCCTGA
GCAAGGAAGCATGAACTGTCTTCACTCTGTCGAAGCATTCCAGTATCAATCCAGCTGCCACTTTAG
TTGTGAAGAGGGATTTGCATTAGTTGGGCCAGAGGTGGTACAGTGCACGGCCTCAGGGATGTGGA
CGGCTGCAGCCCCAGTGTGTGAAGGTGAGTTTCCAGCAGCTGAGGGTGTGTAATGTTCACTGCACC
CAGTTCTTGCTGTGAGTACAGATGACAGTGGCC

Temps: 60, 63.8
Fwd: AAGGGCTCATGCTATTGCTC
Rev: GGCCACTGTCATCTGTACTCAC
Product size: 296bp

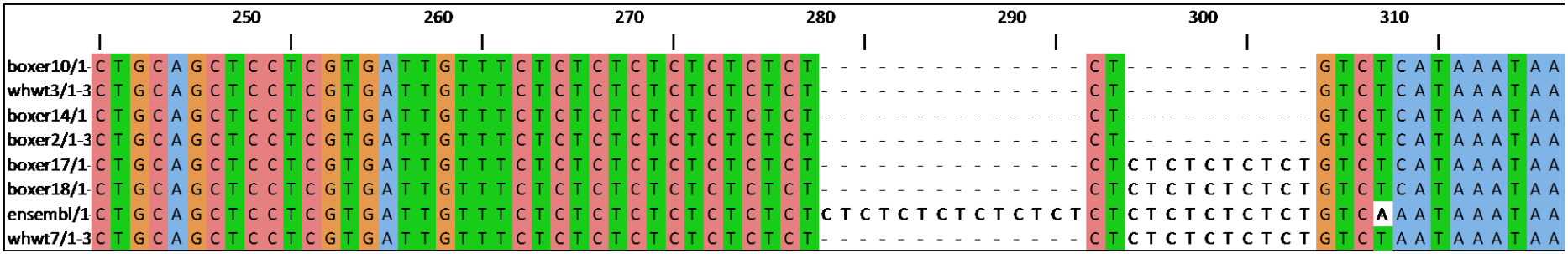
P-selectin 11th exon

GACTCTGTTATGCTCCGTTTATTGGGAGCTCTCTAAATGGAATCATGTTTTCACAGCAGGCATAGTG
TCAGCTCCTACTTCAAAGGTTCAGTGTCCAGCCCTCATCACTCCAGAGCAAGGAACAGTGTCCTGT
AGGCACCATCTGGGAACCTTTGGTCTGAATACCACTTGCTACTTTGGATGCAAAGCTGGATTCATA
CTCATGGGAGACAGTGCTCTCAGATGCAGACCTTCAGGAAAATGGACAGCAGGAACGCCAACATG
CCAAGGTAATGGGAAAGGAACAGGC

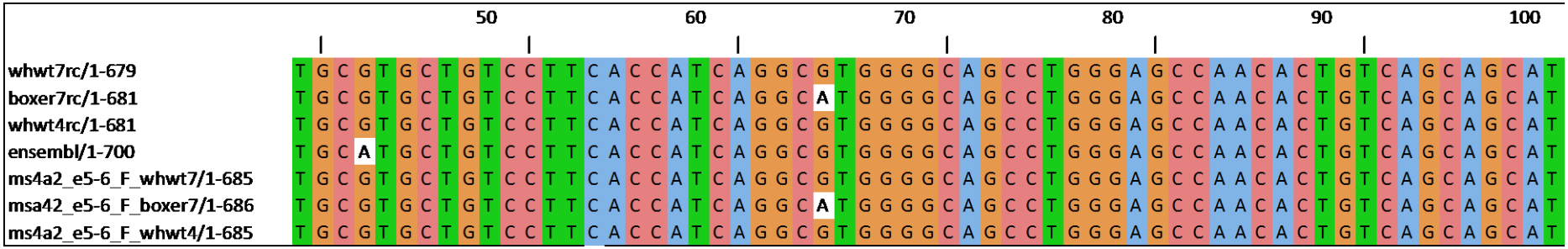
Temp: 60
Fwd: GACTCTGTTATGCTCCGTTTATTG
Rev: GCCTGTTCCTTTCCCATTAC
Product size: 289bp

Appendix 4 Sequences from Jalview

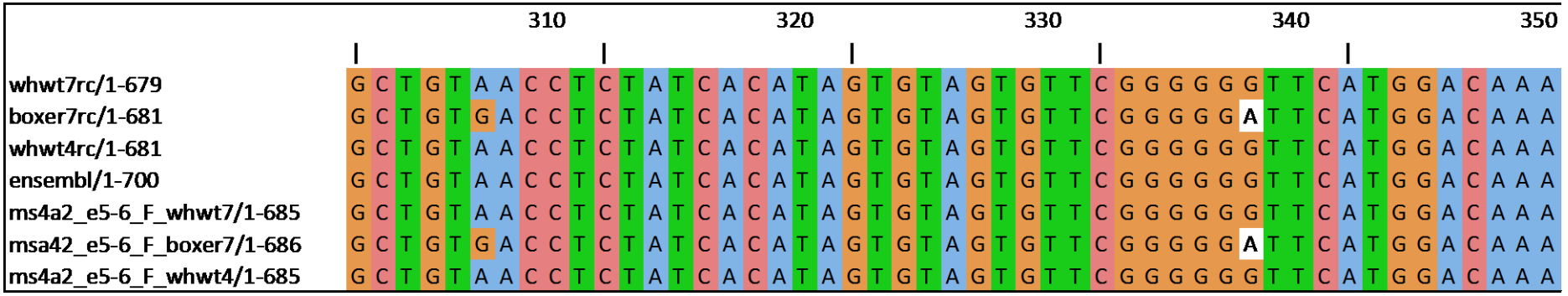
CMA1 promoter region



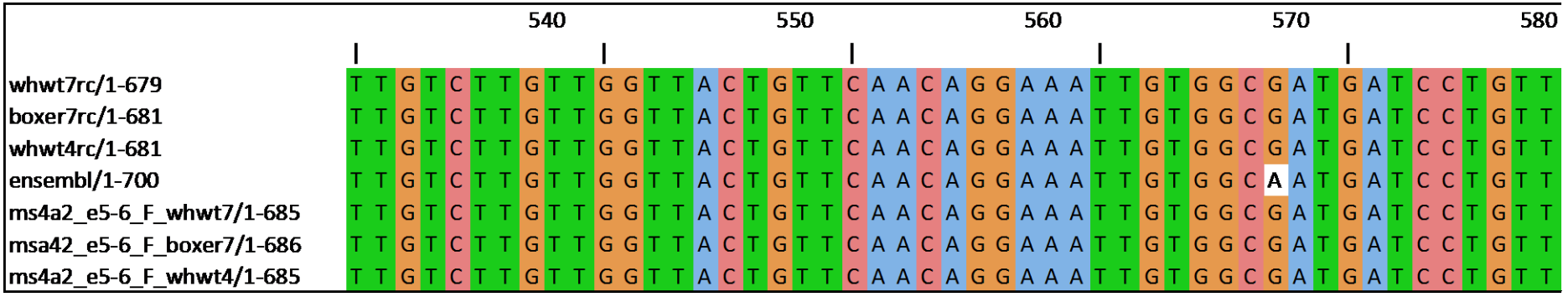
MS4A2 Exon 5-6



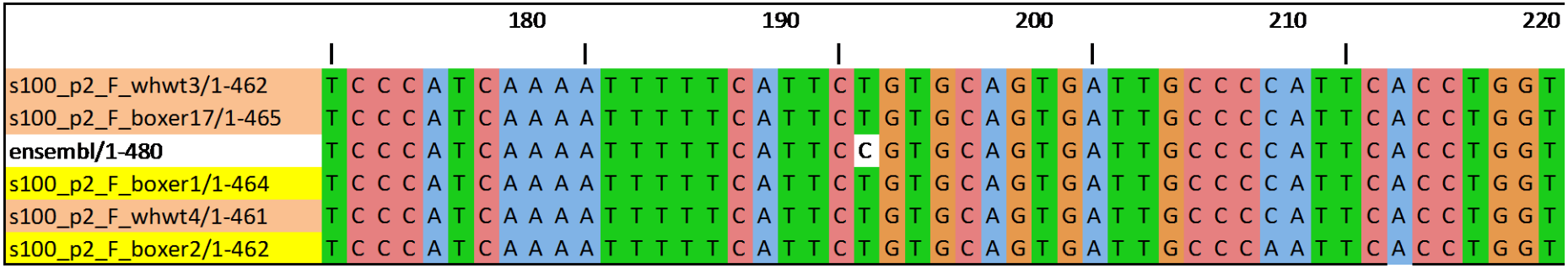
MS4A2 Exon 5-6



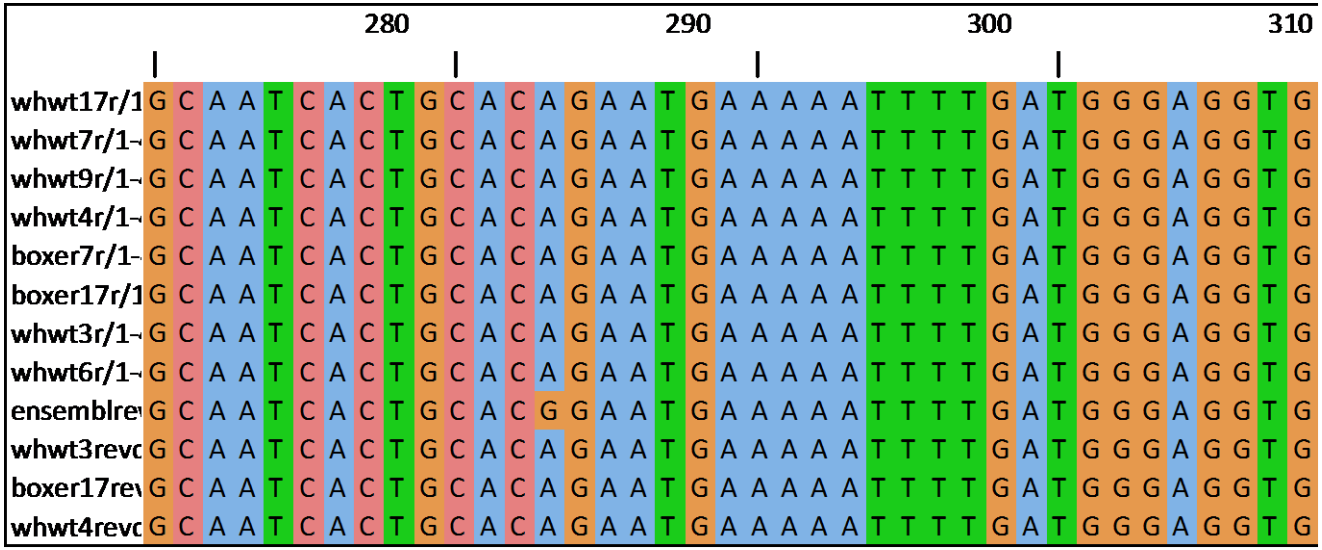
MS4A2 Exon 5-6



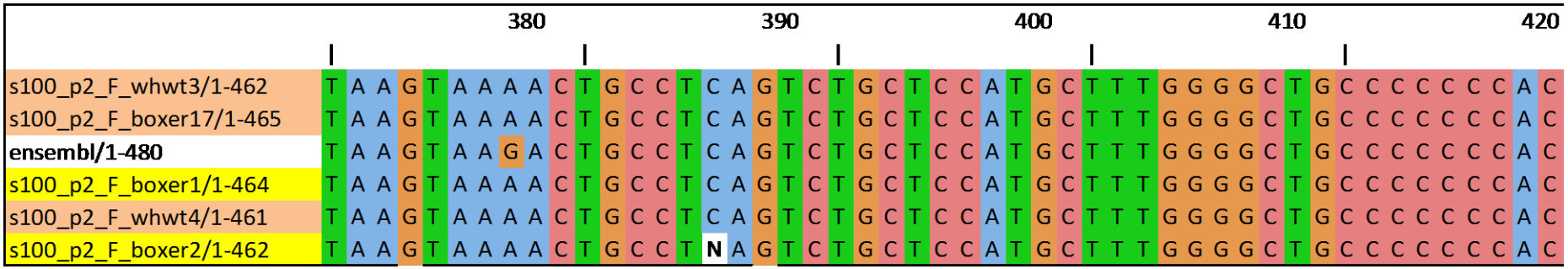
S100A8 promoter



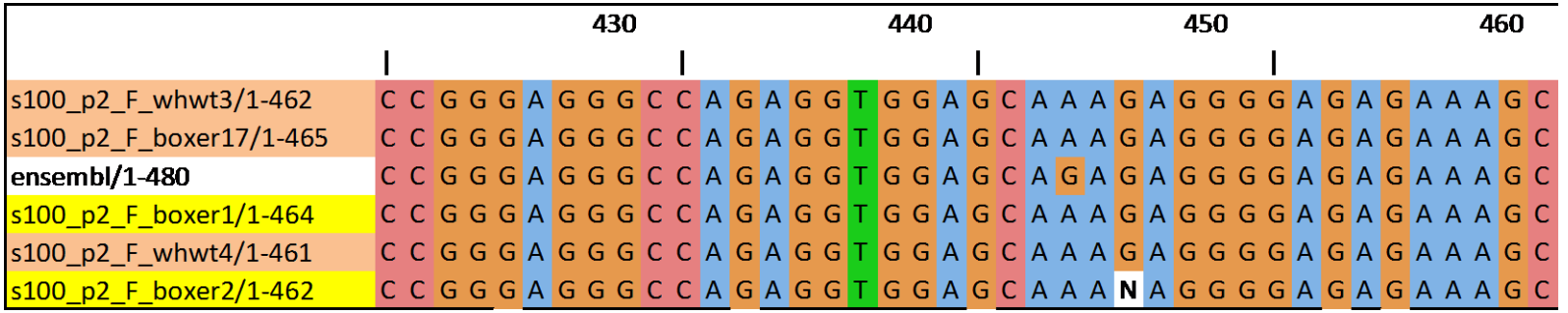
S100A8 promoter



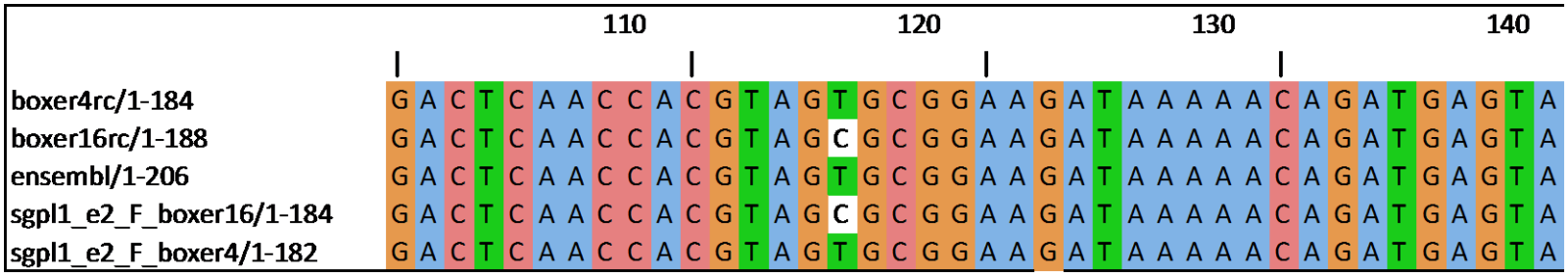
S100A8 promoter



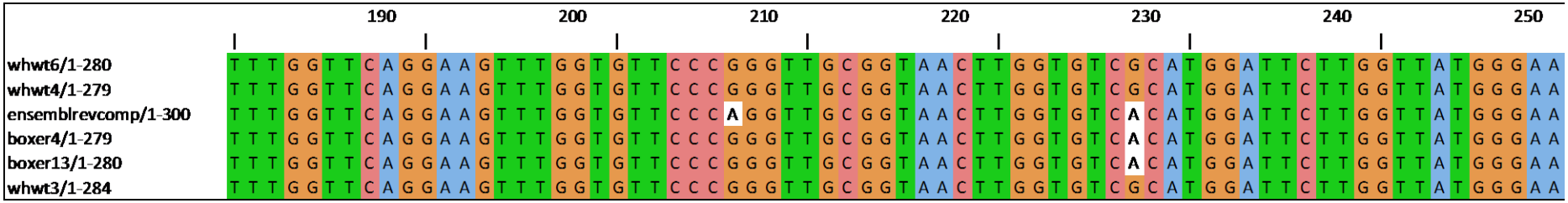
S100A8 promoter



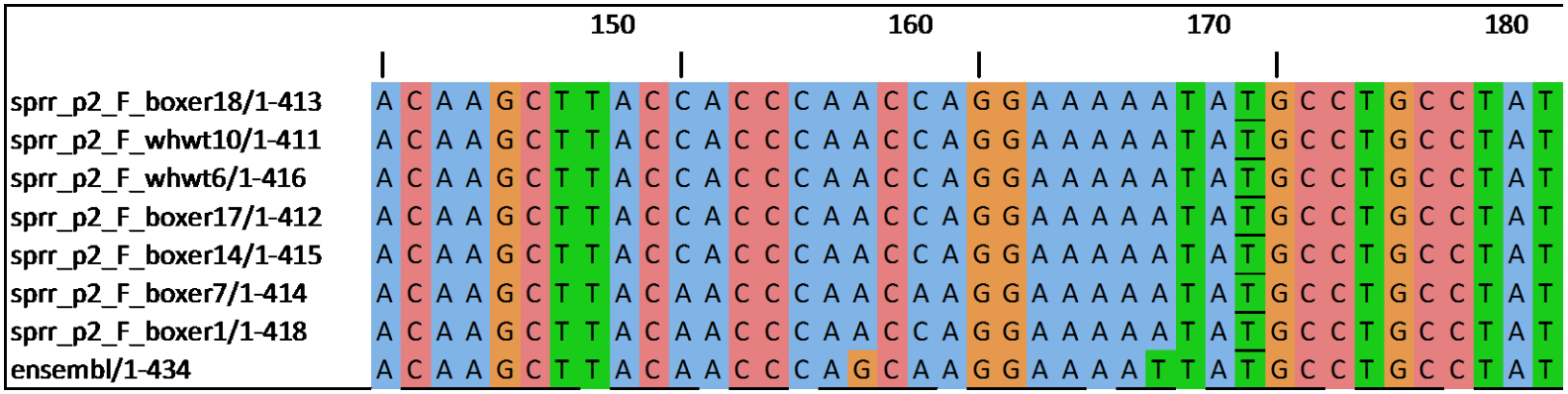
SGPL1 Exon 2



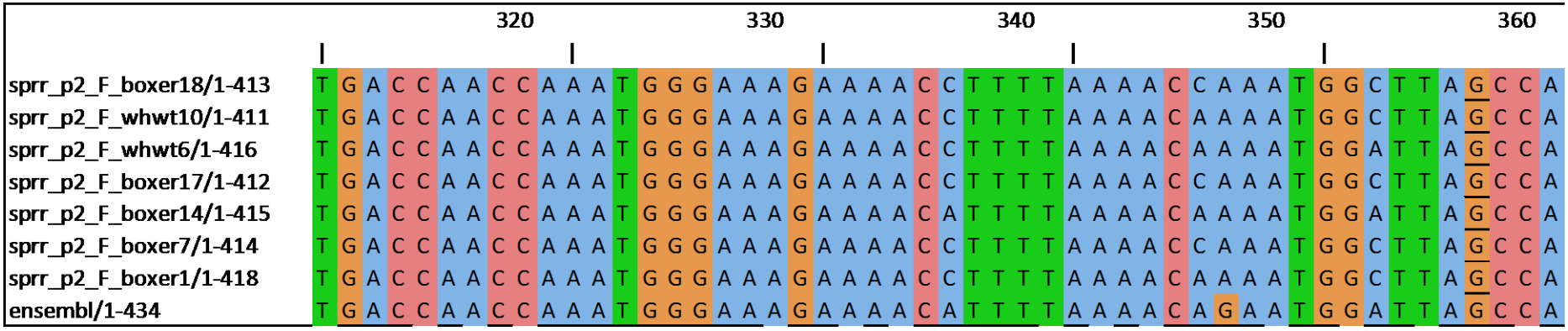
SPRR Exon 3



SPRR1B promoter



SPRR promoter



Appendix 5 Sequences of SNPs found by WAVE

Microsatellites are highlighted in blue. Novel SNPs found on wave are yellow. SNPs already on electronic databases such as ensembl are green. **Peach** and bold represent exons. Pink represents SNPs which were found on WAVE but were already available on electronic databases such as ensembl.

>CMA1 Promoter region ENSCAFG00000012443 (1500bp upstream)

CAGATTCCCCCAGTCACAAGCTCAAGCTAAGACCAAGGGCATGGTAGGATGATTTAGGGAAGGTATGTTTTCCCTGTGATTCCTTATCTCTGAATTAAATGGCT
TCTGAATCTCCAAAAGAGTGGCGCCTGGGTGGCTCAGTCAGTTAAGCATCTGCCTTTGACTCAGGTCATGATCCTGGGGTTCTGGGAGAAAGCTCCATGTCAG
GCTCCTTGTCAGCATGGAATCTGCTTCTCCCTCTTCCTCTGCAGCTCCTCGTGATTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGTCWAAT
AAATAAATAAAATATAATAAATCTTCAAGAGAAACGACACTCTGCCTGTT

>MS4A2 Exon 5-6

ACACACAGGGCTATTGAACGCACTGGCAGGTGGGGGCTCAGGCCTGCRTGCTGTCCTTCACCATCAG**GCRTGGGGCAGCCTGGGAGCCAACACTGTCAGCAG**
CATAGCTGCGGGAATAGGAATCTTTATCCTGACGGTCAACCTGCAGAGGAGCTCGGCTTACATCCACAACTGCCAGCAAGCTCCCAGAGACGACTTCTGCTT
TGTGGCTTGTTTTTCCACAGTACGTATTTCTTGTGAGGGGGCTAAGGTCTGGGTCCTACTCTAAGTACAAGCCCTCTCCTTTCCCACGAACGTCGTCCTTTGCTGT
RACCTCTATCACATAGTGTAGTGTTCGGGGGRTTCATGGACAAAGATGCTGGCTGCCGCTCCCTCCCCAGCTGGGTTCTGTGAGGCTGGAGTCACCCGCCTTGA
CCTCCCTTTTTGTGGCTTTCCCACGTATGTGTCACTTCATTGGGAGTGATCACATCGGGGTGATGAAATCAGGGGCTTCTGTTAGGAACACACCTTTGGGGCACC
TGCCTTGTGACTCTACTTGTCTTGTTGGTTACTGTTCAACAG**GAAATTGTGGCRATGATCCTGTTTCTCACTATTCTGGGGTTTTGCAGTGCTGTGTCACTGACG**
ATCTATGGAGTTGGAGAATTAGTCCAAGGGAACAAGGTAAGTAGAGGCCTGATAATCCTGAATGACAAGCTGAAGC

>S100A8 promoter region

GGGATTAGCAGCAGCAAGAGAGGAGGACTTGAGAGTGGGCATCCAGGCTGGCTGAGCCAAATGATGCCAGTGGCCCAGCAAAGGAGTCCTCACCATGGCCT
CTTCCTGATTTATCCTGCTGGGCACTACTGCCTAGCAAATACCTTCCTCTTTCCATTTTTCCTACTTCCCCACCTCCCATCAAAATTTTTCATTCYGTGCAGTGATTG
CCCCATTCACCTGGTTGARAAACCAGAGGCTGTGGCAACCTGGGGAGGGAGGGGCTATCTCTGCATGGCCTAGGGCCCTGGCAGCTGGCCACAGCCTGGCTG

CTATAAAGGGGAGCTGCCACTGAGCCCTGCATATCCTCTGTCAACTCTGTTTCGGGAAACCTGGTAAGTAARACTGCCTCAGTCTGCTCCATGCTTTGGGGCTG
CCCCCCCACCCCGGGAGGGCCAGAGGTGGAGCARAGAGGGGAGAGAAAGCTAAGGGAGGTCTGGA

>SGPL1 Exon 2

CATTGAAGAAGTGAGTGAAGAACAGACTACCATCAGATGAACTTGTTTTGTACTGGACA**CTGTGGGGGCTTGAAATGATGACTAAG**GTCTGATCAGTGTTCAA
GGACTCAACCACGTAGYGCGGAAGATAAAAACAGATGAGTAGGCTGACTGGTGATAGGCTCTGTGTGGTGGCACAAGGAGATACTTCAGGCGAGTTTAATGG
GAGAGATGA

>SPRR1B exon 3

CAACAGGTGAAACAGACTTGCCAGCCTCTACCTCAG**GAGCCATTTGTTCCCATAACCAAGAATCCATGYGACACCAAGTTACCGCAACCYGGGAACACCAAAC
TTCCTGAACCAAACTACCCTAAATATCCTGAGCCAGGCCATCCCATAATTCCTGAATCAGGCCACATCAAGGTCCCCGAGCCAGCCTACTCTGAGGTCCTTCA
GCCTGGCTACACCAAGGTACCAGAACCAAGTCCCTCAACCATTATTCCAGGCCAGAAGACCAAGCAGAAATGA**TGTGGTCCTTAGCCATGACCT

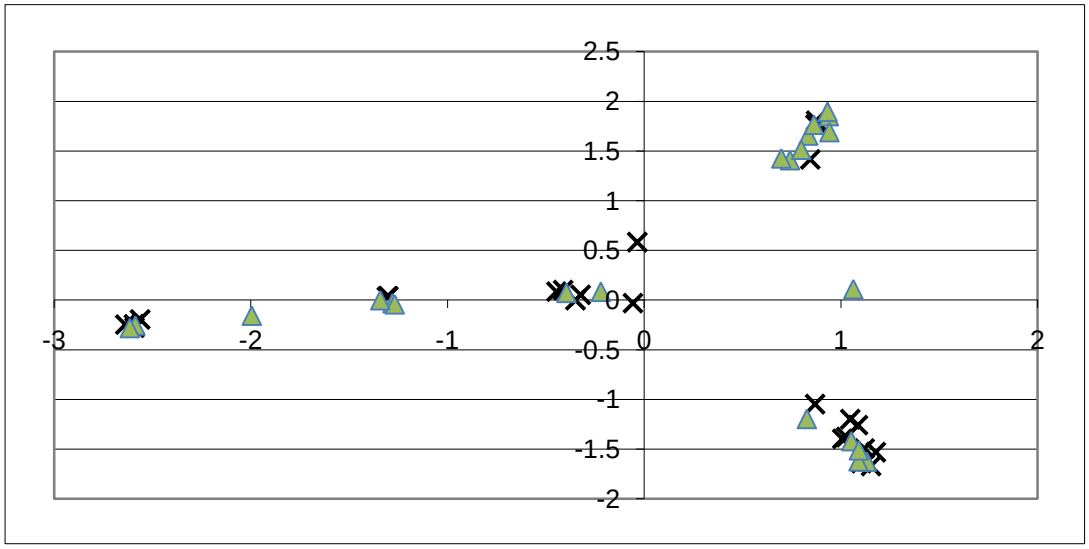
>SPRR1B promoter region

CCCTTTCAGCAACACCACTCTTTTCTCTGTGTATGACACCTATAAAAATGTTGATTCATGAGATGGACCTCTGTTGTGCATTTGTTATATCTCAGGCACCATGCTA
GTCACTTTGGGTAGATCAGACCATTATGGAACTTACTCAGCAAGCTCACAGCCTARCAGGGAAGATAATGCATGCTTATACATAGGTGAAATACAAAGCAGAT
ACTGACTGGTATGGGAACAGAGGGACAGAGAAGGGCCATGGCAGTTTGGGGGAGCTATGCTCAGGATCTCCTGGTACAAGATATGAGCTATAATGGAGGCA
TATCTGGACAAGCAAAATGGGAAGGAGAGCATTTCAGAGCARATTGCATAAGCAAAGGCAGAATGAGGTCACAGGAGCTGAGACTCACAGAGGCATATTTAA
GTCACAGGTATGCCACATGGTC

Appendix 6 Allele size and frequency for microsatellite analysis

Allele	Frequency	Number	Size
A01	0.01	16	150
A03	0.00	4	154
A04	0.02	27	156
A05	0.00	2	158
A06	0.71	842	160
A07	0.03	36	162
A08	0.01	7	164
A09	0.00	3	166
A10	0.12	143	168.5
A11	0.01	12	170.5
A12	0.03	35	172.5
A13	0.05	61	175

Appendix 7 MDS plot of population structure – grouped according to country of origin



USA golden retrievers are represented by green triangles and Japan Golden retrivers are represented by black crosses, the figure represents the intra-breed variation within golden retrievers from USA and Japan, as defined by IBS clustering using genome wide data.

Appendix 8 Primer and probe sequences for GWAS validation analysis 1 (population structure unadjusted) and 2 (population structure adjusted)

SNP	Reverse Primer	Forward Primer	Probe Sequence	Analysis
rs21994080	ACG TTG GAT GAA GAA GGG AGA TCC ACT GGG	ACG TTG GAT GAC ATT GTC TAA TGT TAC TG	GGG GAG ACA TGT CTG CTT AT	2
rs22524918	ACG TTG GAT GTG CTC TCT CTT CCC GTT CAG	ACG TTG GAT GAG GCT TCT GTT GGA AAG CAG	CCC GTT CAG TTC CAA TT	2
rs22576724	ACG TTG GAT GTG TCT ACA GAT GAC TCC CAC	ACG TTG GAT GAA AGG TGA AGC TGA GTC TGG	CTC CCA CTC ATG CTC TTT CAA CA	2
rs22720467	ACG TTG GAT GAG CAA GGC AGG GAT TGT GTC	ACG TTG GAT GAA TGA ATC AGG CAG TCG CTC	TGT GTC TGC AGG CTC TGG CAG CTC G	2
rs22738169	ACG TTG GAT GGT GTG CCA TAT GAG TAG CAG	ACG TTG GAT GCT CCT AGT TAA AAG CAA ATC C	AAG CTG TGA CTT ATC CTC CAT	2
rs22808564	ACG TTG GAT GCT GGA TGC CCT TCC AGT TAA	ACG TTG GAT GTC ACA AGC CCA GGA AAA TTG	GTT AAA AAC TTA TCA ACA ATT CAA TT	2
rs22864357	ACG TTG GAT GAC TGA GCG ACA CCA CTT CAC	ACG TTG GAT GAG ATC TCG CAC GTG TAC TGG	CCG GCA ACC AGC TGC TTA TCG TTC AGA A	2
rs22991453	ACG TTG GAT GGT CCC AGA CTA TGA ACT TGC	ACG TTG GAT GGT TGC TAT TCA CCT CCT CAG	TCC CAG ACT ATG AAC TTG CAT GAT GTT T	2
rs23029421	ACG TTG GAT GGC ATA ATA GAA CTG TGA AGC	ACG TTG GAT GGT ATA TAT AAT TCT GAA AAA TG	AAC ATA GTA AAA TAA CAC GTG ATT AA	2
rs23117702	ACG TTG GAT GCG GTC TGG TCG TTT ATG ATG	ACG TTG GAT GGT AAA TAC ACC AGA AGC CCG	TGC TGC CCC CTT CAC C	2
rs23224877	ACG TTG GAT GAG AAG CTC AAA AGT CAG CCC	ACG TTG GAT GGG AGG AAG TAA TAG TCG GTG	CGT AGG CCT GTT TTC TCC C	2
rs23313105	ACG TTG GAT GGC ACC ACC TCA TTG TTT CTC	ACG TTG GAT GAG CAG GTG CTC AGC AAA TAC	TCC TGC ATG CCC CTT	2
rs23316823	ACG TTG GAT GAC ATA GTG CCT TTG GGA CCG	ACG TTG GAT GTC AGA AAC CCA AAG CAG GAG	TGT GAA CCC CGT GAA	2
rs23444049	ACG TTG GAT GGT AAA ATA GCA AAT CAG TAC	ACG TTG GAT GTT GCA CTA CTG TTA TTA CC	AGC AAA TCA GTA CCA AAA A	2
rs23622647	ACG TTG GAT GAC ATT CAT GGA AAT GTG GC	ACG TTG GAT GGC TCT CTC TCT TTC TCT CTC	ATG GAA ATG TGG CAT TCT GTC C	2
rs23684720	ACG TTG GAT GGG TGA GTG ACT TGC CAT TTG	ACG TTG GAT GGT GTG ATA GCC TGC CTT ATG	GTG ACT TGC CAT TTG AAG TTT GCT	2
rs23708599	ACG TTG GAT GCA TGG GAT ACC CCC TAT GTG	ACG TTG GAT GCT ACT CCC CTC TAA CCT GAC	ACC CCC TAT GTG TTC AGA	2
rs23767031	ACG TTG GAT GCC AAG TAG GGT GGA TTT CTG	ACG TTG GAT GCA GGA AAG CTA ATG GAA CTC	AGG GTG GAT TTC TGG TCC TTT G	2

rs23769251	ACG TTG GAT GCA ATC AAT ACT GCT TCA GGC	ACG TTG GAT GAG AGG CAA TGA CAT ATT GGC	CAT TTC TTC TAC CAT TCA GT	2
rs23769439	ACG TTG GAT GAC TGA CAA GGT GAT CAG AGC	ACG TTG GAT GTG CTT ACC TAT CAG CCA GTC	GGT GAT CAG AGC TAA AAT GA	2
rs23770102	ACG TTG GAT GTA TCT ATG CTT CAG CCT GAC	ACG TTG GAT GGG CAG TGA CGC TTG TAT ATG	AGC CTG ACA GCA GAT GCT AAA	2
rs23803267	ACG TTG GAT GTC CAG TGT TAA AGA ACA GAG	ACG TTG GAT GAG TTG GAA AGA GTT CAG TGG	AGT GTT AAA GAA CAG AGA AGT TAT A	2
rs23856926	ACG TTG GAT GCT TTA TAA AAC AGA GAT GTG G	ACG TTG GAT GAG GAG ACA TTG GTT TTT CCC	ACA GAG ATG TGG AGA AAA	2
rs23865825	ACG TTG GAT GGA ATC TGT TGG CAA CTT GGC	ACG TTG GAT GAT GGT ACG CTA CCT GGA AAG	AAC TTG GCA GGC AGG GAG ATG	2
rs23870923	ACG TTG GAT GGA AAT CTA GGG AGA GGA AGC	ACG TTG GAT GTG GAG AGA CTC TGA TGT CAC	GAA ACA CAT GAA TCC ATA TCA	2
rs23932942	ACG TTG GAT GGT CAC CCA TAT TCT GAC TCC	ACG TTG GAT GTC CAG ATT CTT CAC CCT GAG	GTC TTG GCT TCT CTG TGA	2
rs23968934	ACG TTG GAT GCT TCT ACA GTC AGC CCT TGC	ACG TTG GAT GTT GCG TCG TGA GCT TTG AGG	CCT TGC ACC TAC CTT CGT GGT GTG GAG A	2
rs23974741	ACG TTG GAT GCA CAG TGT GTA AAC TGA GGG	ACG TTG GAT GTA TTT GAG AGT TGG TGG CCC	GCT ACC GTA ATT TTT ATC ACC AT	2
rs24092655	ACG TTG GAT GAG TAG GAG ACT CCT GGG AC	ACG TTG GAT GTG GCC ATG AAG ATG CAT GTG	TCC TGG GAC CCC TCT TC	2
rs24131580	ACG TTG GAT GCC AAT CTC AGA AGG ACT GTG	ACG TTG GAT GGG GAT CAT TTT CAG GTG AGG	GGA CTG TGC ACA TAC AT	2
rs24275059	ACG TTG GAT GCC AAA GCA TTC TTT TCC CAG	ACG TTG GAT GTC GGT AGA ACC ACT GAC TTG	CTT TTG GCC TTT GTA GAT G	2
rs24332727	ACG TTG GAT GTG TGG AGA GAT GGA TGG AAC	ACG TTG GAT GAT CCT TGA GCA ACA GCC TTC	GGT GAG GCT GTG TGT CC	2
rs24354997	ACG TTG GAT GTT GAC TAC TGC ACA CTG CTC	ACG TTG GAT GGT ATG TTA GAG AAG GAA GGG	TCC TCT GTG GAA ACA GT	2
rs24462950	ACG TTG GAT GCT CAT GGC CTA TAT AAG CTC	ACG TTG GAT GAG ACT CAA CTA CGA GCT ATG	ATC AAC AGA GAG CCC T	2
rs24556501	ACG TTG GAT GCG TAG TTC ATA CAT GTT TGG	ACG TTG GAT GAG CTG GGA GTA GAT ATG GTA	CAT GTT TGG ACT AAA GAA ATG AA	2
rs24732893	ACG TTG GAT GTT TCT TTG ATT CTC TGG CGG	ACG TTG GAT GTT GGT ACC AGA TCC AGC TAC	GCG AAG ATC CCC CAA	2
rs9132002	ACG TTG GAT GAA AGT TCC TGG CAA CCA AGC	ACG TTG GAT GAA AAC CAA AGC CCC TTT GCG	CTG GCA AGC TCC CAG	2
rs22020166	ACGTTGGATGGTCTTTC CTTCTTAACCTGC	ACGTTGGATGGGGAATCA AGAAAGGCTTTAC	TATTAACCTGCATA CAAAATTGTTGCT	1

rs22114085	ACGTTGGATGAAGCTCC ACGATGAACAGTC	ACGTTGGATGGACTGTACA CGACCTTTACC	CAGCAGCACTTCC AGAATGTACC	1
rs22155657	ACGTTGGATGATGTTGA AGATTGAATGAC	ACGTTGGATGTATCTAAAC CACCAGTTTC	TGTATATAAAGATT TCTCCTCAA	1
rs22181912	ACGTTGGATGTTCTGCA AGGATGGGCATTC	ACGTTGGATGTCCAGTTTT TCTATGTCCTG	AGAACATGTCCTTG TGG	1
rs22184220	ACGTTGGATGAGACCAT AGCCAGCTTGCAC	ACGTTGGATGTGGCACAAT CCCTTTGATGG	GAGTTAAAGGTAT CTTCCATAGG	1
rs22185491	ACGTTGGATGGGGAAT ACAGCGGTCTTGA	ACGTTGGATGTTGTTCTCC TTCATCCTGGG	GGAATACAGCGGT CTTGACGGGTTG	1
rs22212359	ACGTTGGATGAATCCCT CTTTTCCCCTCTG	ACGTTGGATGGCACAAGA TGGAAGATGGAG	CCCCTCTGAAATTC CCTTT	1
rs22212677	ACGTTGGATGTAACTTC TGCTCAGAAATG	ACGTTGGATGTGTGGTTTT GTATTCAAGGG	AATTGAGAATAAT CAGTGAAATAG	1
rs22328353	ACGTTGGATGACCTGCT ACCAGCATTAAAG	ACGTTGGATGTTGCTTTCC TTCTCCCTTAG	GCTACCAGCATTA AAGCTTTAAC	1
rs22523565	ACGTTGGATGAGACTG GATCTCTTGCTGTG	ACGTTGGATGGAAAAGGA GCTGTGAGGATG	CTTCACAAACTCTT GCC	1
rs22859255	ACGTTGGATGGCATCTC AAATATTTCTGGG	ACGTTGGATGTCAGAGAA AGCCAAAAGCCG	TCTGGGTAAGAGG AACA	1
rs22864357	ACGTTGGATGAGATCTC GCACGTGTACTGG	ACGTTGGATGACTGAGCG ACACCACTTCAC	CGTCCTCCACTACC AC	1
rs22913552	ACGTTGGATGTGAAAC AAGATCACAAAGCC	ACGTTGGATGTGCAGCACT GTGTGTGTTC	CCAAAGCCAAACT GAACAATAATC	1
rs22915894	ACGTTGGATGGCACTAT CGTGTTGTCTAC	ACGTTGGATGCTGAGAATG GCTTCACAGAG	TGTCTACAGAACC AGATTTTGCAGAA GG	1
rs22959786	ACGTTGGATGGGATGC AGACCGTCTGGATT	ACGTTGGATGGTCAGCAG ACAACACTTAAC	CCGTCTGGATTTCC TTA	1
rs22989468	ACGTTGGATGAGGGAG TCACCTGCAAACTG	ACGTTGGATGATGCTCACC TTGCGGAACTC	GCAAATGCAGTTGT GATGCAGTCGTG	1
rs23026134	ACGTTGGATGAGATCCT GGTTTGGAAGGAG	ACGTTGGATGCTTGGTGAG ATCTCCTATGG	GGAGCATTGAGTG AGTCTAT	1
rs23066224	ACGTTGGATGGGGCTAT TAGTGACCCAATG	ACGTTGGATGTCTGCAGTC AGATCAACTGT	CAGAAGTAGGATC ATCTGCTC	1
rs23122280	ACGTTGGATGGATTTCA CTCAGCCCCATAG	ACGTTGGATGAAACAGAT GTCTCTAGTTGG	AGCCCCATAGAGC TTTCTTTAAAT	1
rs23428471	ACGTTGGATGCCACATG CTGATTACAAGCC	ACGTTGGATGGACCACTTT TGTCTAGTAAG	CTATTGGAGTTGTA TTTATCGACAAAC	1
rs23472497	ACGTTGGATGGGCCAG AGGATGGGAAATTC	ACGTTGGATGATCTCTGTC AGTGCACTAGG	ACCTGTCAAGGTC ATGTC	1
rs23478400	ACGTTGGATGTCTGCCT AGAATCTTGTCAC	ACGTTGGATGTTGCCCCCA GACATTTTTTG	GTACTTACATATGC ATTTCCCA	1

Appendix 8 continued

rs23545080	ACGTTGGATGCTTCTAACCTGAGTGAAGGC	ACGTTGGATGTGTGTTTCTGGAGTCCTACC	TGTGTTTGTGGGGTTA	1
rs23567144	ACGTTGGATGGATGAGCCCACACACAGAG	ACGTTGGATGTATCCCTGAAGCATCTGGGC	CCCACACACAGAGGCTGCTGGAATA	1
rs23602938	ACGTTGGATGAGAGGGTCCTGTTGATACAC	ACGTTGGATGGCCAGTAAGCTAGAAGTTCG	CCCATACTCCTCAGGC	1
rs23690306	ACGTTGGATGTACTTCACCTGCAAGCTTGG	ACGTTGGATGGGTAGTTGAGGTAGATGAGG	AATAAGCCAATCCTCCTG	1
rs23768387	ACGTTGGATGTCTCACTTCTCAGTCAGTGC	ACGTTGGATGGAAGAAAGAGATGAGGAGAG	TCACTTCTCAGTCAGTGCTGTCTCATA	1
rs23828846	ACGTTGGATGACTGCAGACTTCCTATCTGG	ACGTTGGATGAGGTCAATCAGACAGTGGTG	TCTGCCCCTCGACAA	1
rs24194054	ACGTTGGATGCAAACAGATTGCTCCGTGTG	ACGTTGGATGGGTACAGAGACCAAGGAAAG	GTTGATGGAGAACAGCATGT	1
rs24267550	ACGTTGGATGTCAATTTGAGCCCGGACTTG	ACGTTGGATGTCAAGAATGGACCCTCTGAC	CGGGTAGATTCTCATAGTC	1
rs24318716	ACGTTGGATGGTTGAACAGGACACCTCATC	ACGTTGGATGACTGTTGGAGGAGGCAGGAG	GAATCGGGCCCATTTGTTTCTA	1
rs24327271	ACGTTGGATGAGAACTCACTCTCTGTGGAC	ACGTTGGATGTGCCAAGAACAAGCCCAAAC	TCTGTGGACTAACTGAAA	1
rs24332727	ACGTTGGATGTGTGGAGAGATGGATGGAAC	ACGTTGGATGATCCTTGAGCAACAGCCTTC	GTGAGGCTGTGTGTCC	1
rs24408651	ACGTTGGATGTCCTCTACCAAATCTATGTG	ACGTTGGATGGCCAACAGTATGCTGGTTAG	CAAATCTATGTGAATAGCCCT	1
rs24482628	ACGTTGGATGCAGGATGCAGATTTTTCCAG	ACGTTGGATGACTGGTGATTCTATGGAGG	TTCCAGTTTTCCCATAGCCTG	1
rs24732893	ACGTTGGATGTTTCTTTGATTCTCTGGCGG	ACGTTGGATGTTGGTACCAGATCCAGCTAC	GGCCGGCGAAGATCCCCCAA	1
rs24872415	ACGTTGGATGTTGGCAAGCCTGACAAATGG	ACGTTGGATGTGTTGAGGAGACTTCACCTG	ACCCAAGTGCCCAAACAA	1
rs8649732	ACGTTGGATGGCTATTCAACTTTGGTGGCG	ACGTTGGATGTGGTTGTTGTCACTAGTTGG	CCTCACAAATCGGCTTAACTCT	1
rs8806978	ACGTTGGATGGATGTGACAAGCCTTTCTGC	ACGTTGGATGAGTCCGTGCTTTATCAAGCC	AGAGGCACCACATGTGCCCAGTTA	1
rs9081246	ACGTTGGATGGACCCAAAGATATTCTGACC	ACGTTGGATGCCGCCCATGTCTCATGAATA	AAAGATATTCTGACCTAAAATACTATG	1

Appendix 9 GWAS Validation Study: Meta-analysis (p <0.05)

Chromosome	SNP	A1	A2	Base Pairs	CHISQ	P	corrected p value (x38)	OR	L95	U95
10	rs22114085	T	C	14087517	21.4	3.7×10^{-6}	0.0001	2.0	1.5	2.7
29	rs23472497	A	G	34162404	16.9	3.9×10^{-5}	0.0015	0.6	0.5	0.8
32	rs23769439	G	A	10826789	8.2	0.0042	0.1598	0.7	0.5	0.9
30	rs23622647	T	C	37858940	7.7	0.0054	0.2070	0.6	0.4	0.9
9	rs24556501	C	T	44666435	7.5	0.0062	0.2360	0.7	0.5	0.9
6	rs24332727	A	G	16258930	7.1	0.0076	0.2875	1.4	1.1	1.8
3	rs23545080	C	A	69975205	7.1	0.0078	0.2963	1.5	1.1	2.0
16	rs22523565	T	C	12801246	6.9	0.0084	0.3208	1.4	1.1	1.8
12	rs22181912	C	T	35696615	6.7	0.0095	0.3596	1.4	1.1	1.8
31	rs23684720	T	G	35018207	6.6	0.0102	0.3880	1.4	1.1	1.8
20	rs22915894	T	C	5771454	6.4	0.0117	0.4438	1.4	1.1	1.7
12	rs22184220	A	G	13111985	6.1	0.0136	0.5179	1.4	1.1	1.7
32	rs23769251	T	C	10876880	5.7	0.0165	0.6281	0.7	0.5	0.9
20	rs8806978	T	C	5779740	5.0	0.0257	0.9774	1.3	1.0	1.7
6	rs24327271	A	C	8499991	5.0	0.0259	0.9857	0.7	0.6	1.0
32	rs23768387	C	T	30365964	4.8	0.0290	1.1012	0.7	0.5	1.0
3	rs23602938	C	G	67125807	4.5	0.0330	1.2544	1.3	1.0	1.8
32	rs23767031	G	T	11760674	3.9	0.0485	1.8445	1.4	1.0	2.0

Appendix 10 GWAS Validation Study: Individual Breed Meta-analysis (p<0.05)

Breed	SNP	Chr	A1	A2	BP	CHISQ	P value	Corrected P value (x48)	ORX	L95	U95
All golden retrievers	RS22859255	2	T	C	48522803	18.1	2.1×10^{-5}	0.0008	5.2	2.4	11.4
All golden retrievers	RS24327271	6	A	C	8499991	13.3	0.0003	0.0101	0.2	0.1	0.5
All golden retrievers	RS23602938	3	C	G	67125807	12.2	0.0005	0.0180	3.9	1.8	8.7
All golden retrievers	RS24408651	7	G	A	7829497	12.1	0.0005	0.0195	0.3	0.1	0.6
All golden retrievers	RS23472497	29	A	G	34162404	11.1	0.0009	0.0329	0.3	0.2	0.6
All golden retrievers	RS23545080	3	C	A	69975205	10.5	0.0012	0.0452	3.7	1.6	8.4
All golden retrievers	RS23622647	30	T	C	37858940	10.3	0.0013	0.0507	0.1	0.0	0.5
All golden retrievers	RS24482628	8	C	A	27285603	9.8	0.0017	0.0652	3.0	1.5	6.0
All golden retrievers	RS22212677	12	A	G	29078487	8.9	0.0028	0.1074	3.3	1.5	7.5
All golden retrievers	RS22523565	16	T	C	12801246	8.3	0.0039	0.1498	2.7	1.4	5.3
All golden retrievers	RS23708599	31	A	G	30830928	7.9	0.0048	0.1834	0.4	0.2	0.8
All golden retrievers	RS23313105	26	C	T	23047328	7.0	0.0080	0.3050	2.7	1.3	5.6
All golden retrievers	RS23690306	31	C	G	25600053	7.0	0.0082	0.3131	2.6	1.3	5.1
All golden retrievers	RS22915894	20	C	T	5771454	6.1	0.0134	0.5092	0.4	0.2	0.8

Appendix 10 continued

All golden retrievers	RS23932942	35	G	A	15274395	5.5	0.0187	0.7095	2.3	1.2	4.4
All golden retrievers	RS23478400	28	C	G	3406416	5.3	0.0212	0.8045	2.2	1.1	4.3
All golden retrievers	RS8806978	20	C	T	5779740	5.0	0.0253	0.9599	0.5	0.2	0.9
All golden retrievers	RS22212359	12	C	G	43043338	4.6	0.0320	1.2145	2.5	1.1	6.1
All golden retrievers	RS23428471	28	G	A	3366214	4.5	0.0333	1.2639	2.1	1.1	4.1
All golden retrievers	RS22184220	12	G	A	13111985	4.0	0.0447	1.6971	2.0	1.0	3.8
All golden retrievers	RS23767031	32	G	T	11760674	4.0	0.0452	1.7168	2.5	1.0	6.0
All golden retrievers	RS24131580	4	A	G	20849589	4.0	0.0455	1.7305	2.7	1.0	7.5
All Gsd	RS22184220	12	A	G	13111985	7.7	0.0054	0.2054	5.3	1.7	16.7
All Gsd	RS24318716	6	C	T	4224712	6.2	0.0127	0.4834	0.1	0.0	0.8
All Gsd	RS23472497	29	G	A	34162404	5.8	0.0159	0.6046	3.2	1.2	8.6
All Gsd	RS23026134	22	T	C	55676006	5.4	0.0201	0.7634	3.7	1.2	11.7
All Gsd	RS23313105	26	C	T	23047328	4.3	0.0375	1.4242	3.0	1.0	8.9
All Gsd	RS22720467	19	C	T	32273778	3.8	0.0508	1.9285	0.0	NA	NA
All Labs	RS24354997	6	T	G	14408543	7.2	0.0071	0.2712	2.4	1.3	4.5
All Labs	RS23708599	31	A	G	30830928	6.7	0.0097	0.3692	2.0	1.2	3.4
All Labs	RS23026134	22	T	C	55676006	5.0	0.0259	0.9853	0.5	0.2	1.0

Appendix 10 continued

All Labs	RS24275059	6	C	A	14365278	3.9	0.0478	1.8156	1.8	1.0	3.2
All Whwt	RS24482628	8	A	C	27285603	11.7	0.0006	0.0232	8.1	2.2	29.7
All Whwt	RS22915894	20	C	T	5771454	11.1	0.0009	0.0337	5.4	1.8	15.7
All Whwt	RS8806978	20	C	T	5779740	11.1	0.0009	0.0337	5.4	1.8	15.7
All Whwt	RS23224877	25	A	G	29485983	8.1	0.0045	0.1722	6.3	1.4	27.4
All Whwt	RS24194054	5	T	C	56199928	7.7	0.0056	0.2132	4.3	1.4	13.3
All Whwt	RS23769439	32	G	A	10826789	4.4	0.0368	1.3995	3.1	1.1	9.2
All Whwt	RS23769251	32	T	C	10876880	3.9	0.0487	1.8502	2.9	1.0	8.4
All Whwt	RS23684720	31	T	G	35018207	3.8	0.0508	1.9304	3.5	0.9	12.8
All Whwt	RS22738169	2	T	C	45023367	3.8	0.0509	1.9338	2.7	1.0	7.4

Appendix 11 GWAS Validation Study: Individual breed analysis (p <0.05)

Breed	Statistical test	SNP	CHR	BP	A1	F_A	F_U	A2	CHISQ	ORX	P	Corrected p value (x38)	L95	U95
Boxer	Allelic	RS22913552	20	33818500	G	0.2	0.5	T	7.4	0.2	0.0065	0.2471	0.0	0.6
Boxer	Allelic	RS23968934	36	32710839	C	0.6	0.3	T	6.0	4.5	0.0145	0.5510	1.3	15.5
Boxer	Allelic	RS23478400	28	3406416	C	0.4	0.1	G	4.2	4.5	0.0411	1.5618	1.0	20.3
Japan golden retriever	Allelic	RS22576724	17	52668444	C	0.1	0.5	T	8.8	0.1	0.0031	0.1172	0.0	0.6
Japan golden retriever	Allelic	RS23545080	3	69975205	C	0.5	0.1	A	8.5	6.3	0.0035	0.1337	1.7	23.7
Japan golden retriever	Allelic	RS22181912	12	35696615	T	0.2	0.6	C	8.3	0.2	0.0040	0.1523	0.1	0.6
Japan golden retriever	Allelic	RS24408651	7	7829497	G	0.2	0.6	A	8.3	0.2	0.0040	0.1523	0.1	0.6
Japan golden retriever	Allelic	RS24872415	3	62441438	A	0.6	0.3	G	7.5	4.6	0.0061	0.2307	1.5	14.3
Japan golden retriever	Allelic	RS22859255	2	48522803	T	0.5	0.2	C	6.9	4.7	0.0087	0.3312	1.4	15.4
Japan golden retriever	Allelic	RS22808564	2	55063211	T	0.4	0.1	C	6.9	5.4	0.0088	0.3331	1.4	20.1
Japan golden retriever	Allelic	RS22913552	20	33818500	G	0.7	0.3	T	6.7	4.2	0.0099	0.3751	1.4	12.7
Japan golden retriever	Allelic	RS23066224	22	39938082	T	0.2	0.0	C	6.7	NA	0.0099	0.3753	NA	NA
Japan golden retriever	Allelic	RS24732893	12	40204012	G	0.0	0.3	A	5.9	0.1	0.0152	0.5765	0.0	0.9
Japan golden retriever	Allelic	RS8649732	19	45495526	A	0.2	0.5	G	5.8	0.2	0.0165	0.6259	0.1	0.8
Japan golden retriever	Allelic	RS23602938	3	67125807	C	0.5	0.2	G	5.7	3.8	0.0168	0.6384	1.2	11.9
Japan golden retriever	Allelic	RS22212677	12	29078487	A	0.3	0.1	G	5.5	5.2	0.0190	0.7228	1.2	22.2
Japan golden retriever	Allelic	RS24332727	6	16258930	G	0.3	0.6	A	5.5	0.3	0.0193	0.7345	0.1	0.8

Appendix 11 continued

Japan golden retriever	Allelic	RS23117702	23	39572738	T	0.1	0.4	C	4.7	0.2	0.0308	1.1700	0.1	0.9
Japan golden retriever	Allelic	RS24131580	4	20849589	A	0.3	0.1	G	4.3	5.3	0.0376	1.4277	1.0	29.2
Japan golden retriever	Allelic	RS23968934	36	32710839	T	0.5	0.2	C	4.2	3.3	0.0407	1.5458	1.0	10.4
Japan golden retriever	Allelic	RS22212359	12	43043338	C	0.3	0.1	G	4.1	4.3	0.0434	1.6484	1.0	18.6
Japan golden retriever	Allelic	RS23974741	36	26538495	A	0.4	0.1	T	4.0	3.5	0.0457	1.7374	1.0	12.3
Japan Labrador	Allelic	RS22808564	2	55063211	T	0.0	0.5	C	9.9	0.0	0.0017	0.0638	0.0	NA
Japan Labrador	Allelic	RS23767031	32	11760674	G	0.3	0.0	T	7.9	NA	0.0049	0.1861	NA	NA
Japan Labrador	Allelic	RS24354997	6	14408543	T	0.4	0.1	G	4.5	6.0	0.0339	1.2893	1.0	35.3
Japan Labrador	Allelic	RS24327271	6	8499991	A	0.4	0.1	C	4.3	4.9	0.0374	1.4212	1.0	23.6
Pit Bull	Allelic	RS22212677	12	29078487	A	0.1	0.4	G	6.9	0.2	0.0088	0.3348	0.1	0.7
Pit Bull	Allelic	RS23769251	32	10876880	T	0.2	0.5	C	6.5	0.3	0.0106	0.4039	0.1	0.7
Pit Bull	Allelic	RS22524918	17	8695250	C	0.2	0.0	T	6.0	9.6	0.0143	0.5445	1.1	80.7
Pit Bull	Allelic	RS24354997	6	14408543	T	0.2	0.4	G	5.4	0.3	0.0199	0.7573	0.1	0.8
Pit Bull	Allelic	RS23865825	34	17995250	T	0.3	0.5	C	5.3	0.3	0.0217	0.8246	0.1	0.9
Pit Bull	Allelic	RS24732893	12	40204012	A	0.6	0.3	G	5.1	3.2	0.0238	0.9052	1.2	9.0
Pit Bull	Allelic	RS23545080	3	69975205	C	0.1	0.3	A	4.7	0.3	0.0308	1.1712	0.1	0.9
Pit Bull	Allelic	RS23313105	26	23047328	C	0.5	0.3	T	4.6	3.0	0.0323	1.2274	1.1	8.3
Pit Bull	Allelic	RS23769439	32	10826789	G	0.3	0.5	A	4.2	0.4	0.0411	1.5603	0.1	1.0
Shiba Inu	Allelic	RS23767031	32	11760674	G	0.3	0.0	T	18.1	26.4	0.0000	0.0008	3.3	211.2
Shiba Inu	geno/trend	RS23767031	32		-	-	-	-	-	-	0.0000	0.0009	-	-
Shiba Inu	Allelic	RS24092655	4	8627345	T	0.4	0.1	C	15.3	7.4	0.0001	0.0035	2.5	22.2
Shiba Inu	logistic/ADD	RS23767031	32		-	-	-	-	-	-	0.0013	0.0508	4.0	297.5
Shiba Inu	Allelic	RS24332727	6	16258930	G	0.3	0.6	A	7.9	0.3	0.0049	0.1847	0.1	0.7
Shiba Inu	Allelic	RS24267550	5	18815805	A	0.2	0.0	G	7.7	5.8	0.0057	0.2153	1.5	22.5

Appendix 11 continued

Shiba Inu	Allelic	RS23622647	30	37858940	T	0.1	0.0	C	7.6	NA	0.0058	0.2187	NA	NA
Shiba Inu	Allelic	RS23313105	26	23047328	C	0.3	0.5	T	7.6	0.3	0.0058	0.2193	0.1	0.7
Shiba Inu	Allelic	RS24275059	6	14365278	C	0.3	0.1	A	7.4	4.4	0.0064	0.2417	1.4	13.7
Shiba Inu	Allelic	RS23968934	36	32710839	T	0.5	0.3	C	7.4	3.0	0.0066	0.2491	1.3	6.8
Shiba Inu	Allelic	RS24872415	3	62441438	G	0.5	0.3	A	7.4	3.0	0.0066	0.2491	1.3	6.8
Shiba Inu	Allelic	RS23870923	34	5326438	C	0.4	0.1	T	7.3	3.5	0.0069	0.2636	1.4	8.9
Shiba Inu	Allelic	RS24354997	6	14408543	T	0.3	0.1	G	7.1	4.3	0.0079	0.3010	1.4	13.2
Shiba Inu	Allelic	RS23865825	34	17995250	T	0.3	0.1	C	6.0	3.4	0.0144	0.5487	1.2	9.4
Shiba Inu	Allelic	RS23974741	36	26538495	A	0.2	0.4	T	5.4	0.3	0.0206	0.7839	0.1	0.9
Shiba Inu	Allelic	RS22212359	12	43043338	C	0.2	0.1	G	5.3	3.7	0.0208	0.7912	1.2	11.7
Shiba Inu	Allelic	RS22991453	21	18223929	G	0.0	0.2	A	5.1	0.1	0.0242	0.9188	0.0	1.0
Shiba Inu	Allelic	RS22524918	17	8695250	C	0.1	0.0	T	4.8	8.1	0.0289	1.0982	0.9	71.7
Shiba Inu	Allelic	RS23828846	33	29024878	T	0.5	0.3	C	4.7	2.4	0.0296	1.1244	1.1	5.4
Shiba Inu	Allelic	RS23066224	22	39938082	T	0.2	0.1	C	4.7	4.2	0.0308	1.1708	1.1	17.0
Shiba Inu	Allelic	RS22931896	20	9213048	A	0.3	0.1	G	4.5	3.1	0.0342	1.2977	1.1	9.2
Shiba Inu	Allelic	RS22576724	17	52668444	C	0.3	0.2	T	4.1	2.5	0.0435	1.6530	1.0	6.1
Shih Tzu	Allelic	RS24092655	4	8627345	T	0.0	0.1	C	8.6	0.1	0.0034	0.1288	0.0	0.7
Shih Tzu	Allelic	RS23865825	34	17995250	T	0.1	0.2	C	7.0	0.3	0.0080	0.3034	0.1	0.7
Shih Tzu	Allelic	RS23708599	31	30830928	A	0.5	0.3	G	5.9	2.0	0.0155	0.5871	1.1	3.5
Shih Tzu	Allelic	RS23026134	22	55676006	T	0.1	0.3	C	5.2	0.4	0.0224	0.8523	0.2	0.9
Shih Tzu	Allelic	RS22808564	2	55063211	T	0.2	0.4	C	5.1	0.5	0.0241	0.9173	0.3	0.9
Shih Tzu	Allelic	RS23472497	29	34162404	A	0.3	0.4	G	4.3	0.5	0.0385	1.4630	0.3	1.0
UK golden retriever	Allelic	RS23545080	3	69975205	C	0.6	0.1	A	31.0	29.3	0.0000	0.0000	5.8	148.2
UK golden retriever	Allelic	RS22915894	20	5771454	T	1.0	0.2	C	30.0	NA	0.0000	0.0000	NA	NA
UK golden retriever	Allelic	RS8806978	20	5779740	T	1.0	0.2	C	29.3	NA	0.0000	0.0000	NA	NA

Appendix 11 continued

UK golden retriever	geno/trend	RS23545080	3		-	-	-	-	-	-	0.0000	0.0000	-	-
UK golden retriever	geno/trend	RS22915894	20		-	-	-	-	-	-	0.0000	0.0000	-	-
UK golden retriever	geno/trend	RS8806978	20		-	-	-	-	-	-	0.0000	0.0000	-	-
UK golden retriever	Allelic	RS23968934	36	32710839	T	1.0	0.3	C	19.4	NA	0.0000	0.0004	NA	NA
UK golden retriever	geno/trend	RS22864357	20	28025372	-	-	-	-	-	-	0.0000	0.0016	-	-
UK golden retriever	Allelic	RS24332727	6	16258930	A	1.0	0.3	G	14.0	NA	0.0002	0.0068	NA	NA
UK golden retriever	Allelic	RS23026134	22	55676006	T	0.5	0.1	C	12.2	9.8	0.0005	0.0180	2.2	44.4
UK golden retriever	Allelic	RS23690306	31	25600053	C	0.8	0.2	G	11.7	11.2	0.0006	0.0238	2.1	58.7
UK golden retriever	geno/trend	RS23690306	31		-	-	-	-	-	-	0.0007	0.0263	-	-
UK golden retriever	geno/trend	RS23026134	22		-	-	-	-	-	-	0.0008	0.0311	-	-
UK golden retriever	Allelic	RS23472497	29	34162404	G	0.9	0.4	A	8.7	12.8	0.0033	0.1241	1.5	107.1
UK golden retriever	Allelic	RS24872415	3	62441438	A	0.9	0.4	G	8.3	12.4	0.0039	0.1473	1.5	103.6
UK golden retriever	Allelic	RS24092655	4	8627345	T	0.8	0.3	C	8.0	7.8	0.0048	0.1811	1.5	40.6
UK golden retriever	Allelic	RS24194054	5	56199928	C	0.6	0.2	T	7.4	6.4	0.0066	0.2522	1.4	28.3
UK golden retriever	Allelic	RS22913552	20	33818500	G	0.8	0.3	T	7.1	7.2	0.0077	0.2922	1.4	37.2
UK golden retriever	Allelic	RS22212677	12	29078487	A	0.3	0.0	G	6.9	8.2	0.0088	0.3346	1.3	51.3
UK golden retriever	Allelic	RS23602938	3	67125807	C	0.6	0.2	G	6.6	5.8	0.0105	0.3971	1.3	25.8

Appendix 11 continued

UK golden retriever	Allelic	RS23974741	36	26538495	A	0.0	0.4	T	5.8	0.0	0.0160	0.6091	0.0	NA
UK golden retriever	Allelic	RS24275059	6	14365278	C	0.0	0.4	A	5.6	0.0	0.0177	0.6718	0.0	NA
UK golden retriever	Allelic	RS22720467	19	32273778	T	0.9	0.5	C	5.4	8.4	0.0206	0.7843	1.0	70.4
UK golden retriever	Allelic	RS24408651	7	7829497	A	0.8	0.4	G	4.6	5.2	0.0313	1.1875	1.0	26.7
UK golden retriever	Allelic	RS24131580	4	20849589	A	0.0	0.4	G	4.5	0.0	0.0333	1.2662	0.0	NA
UK golden retriever	Allelic	RS23708599	31	30830928	A	0.1	0.5	G	4.4	0.1	0.0356	1.3509	0.0	1.2
UK GSD	Allelic	RS22720467	19	32273778	C	0.4	0.0	T	17.6	NA	0.0000	0.0011	NA	NA
UK GSD	Allelic	RS23769251	32	10876880	T	0.3	0.0	C	14.4	NA	0.0002	0.0057	NA	NA
UK GSD	Allelic	RS24872415	3	62441438	G	0.4	0.0	A	13.4	15.6	0.0002	0.0094	2.8	87.7
UK GSD	Allelic	RS24275059	6	14365278	C	0.3	0.0	A	10.4	18.6	0.0013	0.0476	2.0	176.4
UK GSD	Allelic	RS24354997	6	14408543	T	0.3	0.0	G	10.4	18.6	0.0013	0.0476	2.0	176.4
UK GSD	Allelic	RS23316823	26	18288615	A	0.2	0.0	G	8.3	NA	0.0040	0.1503	NA	NA
UK GSD	Allelic	RS24092655	4	8627345	T	0.2	0.0	C	8.3	NA	0.0040	0.1503	NA	NA
UK GSD	Allelic	RS23224877	25	29485983	A	0.3	0.0	G	7.5	13.7	0.0061	0.2312	1.4	134.1
UK GSD	Allelic	RS24732893	12	40204012	A	0.3	0.1	G	7.2	8.6	0.0073	0.2770	1.5	50.8
UK GSD	Allelic	RS24556501	9	44666435	T	0.4	0.1	C	7.2	5.8	0.0075	0.2832	1.5	22.4
UK GSD	Allelic	RS23472497	29	34162404	G	0.7	0.3	A	5.9	4.4	0.0149	0.5670	1.3	15.2
UK GSD	Allelic	RS23690306	31	25600053	G	0.7	0.4	C	5.1	4.0	0.0238	0.9025	1.2	13.6
UK GSD	Allelic	RS24408651	7	7829497	A	0.4	0.1	G	4.9	4.4	0.0263	0.9979	1.1	17.6
UK GSD	Allelic	RS22184220	12	13111985	A	0.3	0.1	G	4.1	4.9	0.0434	1.6481	0.9	25.7
UK Labrador	Allelic	RS22915894	20	5771454	T	0.5	0.2	C	16.8	4.4	0.0000	0.0016	2.1	9.2
UK Labrador	Allelic	RS9132002	29	22197558	G	0.2	0.0	A	16.4	23.2	0.0001	0.0020	2.8	194.0
UK Labrador	geno/trend	RS22114085	10	14087517	-	-	-	-	-	-	0.0001	0.0052	-	-
UK Labrador	Allelic	RS8806978	20	5779740	T	0.5	0.2	C	14.2	3.9	0.0002	0.0064	1.9	8.1

Appendix 11 continued

UK Labrador	geno/trend	RS22915894	20		-	-	-	-	-	-	0.0003	0.0104	-	-
UK Labrador	logistic/ADD	RS22114085	10	14087517	-	-	-	-	-	-	0.0005	0.0208	1.8	7.7
UK Labrador	logistic/ADD	RS22915894	20		-	-	-	-	-	-	0.0008	0.0296	1.7	7.0
UK Labrador	geno/trend	RS8806978	20		-	-	-	-	-	-	0.0008	0.0323	-	-
UK Labrador	Allelic	RS22114085	10	14087517	T	0.6	0.2	C	20.9	6.1	4.8 x 10-6	0.0002	2.7	14.1
UK Labrador	Allelic	RS22720467	19	32273778	C	0.3	0.5	T	11.1	0.3	0.0009	0.0329	0.1	0.6
UK Labrador	geno/trend	RS9132002	29		-	-	-	-	-	-	0.0009	0.0333	-	-
UK Labrador	Allelic	RS24275059	6	14365278	C	0.2	0.4	A	10.9	0.2	0.0009	0.0360	0.1	0.6
UK Labrador	Allelic	RS24327271	6	8499991	A	0.2	0.5	C	9.4	0.3	0.0021	0.0814	0.1	0.7
UK Labrador	Allelic	RS23545080	3	69975205	C	0.2	0.1	A	8.3	4.3	0.0040	0.1534	1.5	12.3
UK Labrador	Allelic	RS24872415	3	62441438	A	0.6	0.4	G	7.1	2.5	0.0078	0.2953	1.3	5.0
UK Labrador	Allelic	RS22212677	12	29078487	A	0.2	0.0	G	6.7	4.4	0.0094	0.3577	1.3	14.7
UK Labrador	Allelic	RS23684720	31	35018207	T	0.3	0.2	G	5.7	2.5	0.0174	0.6616	1.2	5.4
UK Labrador	Allelic	RS22913552	20	33818500	G	0.5	0.3	T	4.8	2.2	0.0291	1.1066	1.1	4.4
UK Labrador	Allelic	RS24194054	5	56199928	C	0.4	0.2	T	4.7	2.2	0.0293	1.1145	1.1	4.7
UK Labrador	Allelic	RS24131580	4	20849589	A	0.2	0.4	G	4.7	0.4	0.0306	1.1628	0.2	0.9
UK Labrador	Allelic	RS24556501	9	44666435	C	0.3	0.5	T	4.1	0.5	0.0423	1.6089	0.2	1.0
UK Labrador	Allelic	RS23767031	32	11760674	G	0.1	0.3	T	4.0	0.4	0.0452	1.7157	0.2	1.0
UK Labrador	Allelic	RS23472497	29	34162404	G	0.5	0.4	A	4.0	2.0	0.0456	1.7309	1.0	3.9
UK WHWT	Allelic	RS24482628	8	27285603	A	0.5	0.1	C	14.4	11.0	0.0002	0.0057	2.8	43.7
UK WHWT	geno/trend	RS24482628	8		-	-	-	-	-	-	0.0003	0.0107	-	-
UK WHWT	Allelic	RS23224877	25	29485983	A	0.2	0.0	G	11.4	NA	0.0008	0.0286	NA	NA
UK WHWT	Allelic	RS24194054	5	56199928	T	0.7	0.3	C	11.2	7.0	0.0008	0.0316	2.1	23.5
UK WHWT	geno/trend	RS24194054	5		-	-	-	-	-	-	0.0011	0.0427	-	-
UK WHWT	Allelic	RS23708599	31	30830928	G	0.4	0.1	A	7.1	5.5	0.0076	0.2874	1.5	20.6
UK WHWT	Allelic	RS23066224	22	39938082	T	0.1	0.0	C	5.5	NA	0.0190	0.7228	NA	NA
UK WHWT	Allelic	RS24872415	3	62441438	G	0.1	0.0	A	5.5	NA	0.0190	0.7228	NA	NA

Appendix 11 continued

UK WHWT	Allelic	RS22523565	16	12801246	T	0.2	0.0	C	4.9	9.4	0.0270	1.0264	0.9	97.3
UK WHWT	Allelic	RS24092655	4	8627345	T	0.2	0.0	C	4.9	9.4	0.0270	1.0264	0.9	97.3
UK WHWT	Allelic	RS23684720	31	35018207	T	0.3	0.1	G	4.2	4.2	0.0404	1.5337	1.0	18.1
USA golden retriever	Allelic	RS23316823	26	18288615	A	0.0	0.4	G	13.9	0.0	0.0002	0.0075	0.0	0.4
USA golden retriever	Allelic	RS23708599	31	30830928	A	0.3	0.7	G	12.8	0.2	0.0003	0.0129	0.1	0.5
USA golden retriever	Allelic	RS24327271	6	8499991	A	0.1	0.5	C	12.6	0.1	0.0004	0.0145	0.0	0.4
USA golden retriever	Allelic	RS22523565	16	12801246	T	0.6	0.2	C	11.9	5.4	0.0006	0.0219	2.0	14.7
USA golden retriever	Allelic	RS23770102	32	11802242	C	0.0	0.3	G	11.4	0.0	0.0007	0.0275	0.0	NA
USA golden retriever	geno/trend	RS23708599	31		-	-	-	-	-	-	0.0011	0.0405	-	-
USA golden retriever	Allelic	RS22859255	2	48522803	T	0.5	0.1	C	10.3	5.7	0.0013	0.0502	1.9	17.6
USA golden retriever	Allelic	RS23472497	29	34162404	A	0.3	0.6	G	10.0	0.2	0.0015	0.0587	0.1	0.6
USA golden retriever	Allelic	RS23684720	31	35018207	T	0.6	0.3	G	10.0	4.5	0.0015	0.0587	1.7	11.7
USA golden retriever	Allelic	RS23622647	30	37858940	T	0.1	0.3	C	7.4	0.1	0.0064	0.2424	0.0	0.7
USA golden retriever	Allelic	RS23313105	26	23047328	C	0.4	0.2	T	7.4	4.2	0.0066	0.2501	1.4	12.2
USA golden retriever	Allelic	RS24482628	8	27285603	C	0.6	0.3	A	7.4	3.6	0.0066	0.2526	1.4	9.1
USA golden retriever	Allelic	RS23602938	3	67125807	C	0.4	0.1	G	7.2	4.8	0.0074	0.2820	1.4	16.4
USA golden retriever	Allelic	RS24408651	7	7829497	A	0.6	0.3	G	6.2	3.2	0.0125	0.4742	1.3	8.2
USA golden retriever	Allelic	RS8649732	19	45495526	A	0.1	0.4	G	6.1	0.2	0.0133	0.5035	0.1	0.8
USA golden	Allelic	RS24318716	6	4224712	C	0.2	0.5	T	5.6	0.3	0.0179	0.6802	0.1	0.8

Appendix 11 continued

USA golden retriever	Allelic	RS23690306	31	25600053	C	0.5	0.3	G	5.1	2.9	0.0244	0.9264	1.1	7.6
USA golden retriever	Allelic	RS22915894	20	5771454	T	0.6	0.4	C	5.0	2.8	0.0252	0.9561	1.1	6.9
USA golden retriever	Allelic	RS23026134	22	55676006	T	0.4	0.2	C	4.9	3.1	0.0262	0.9956	1.1	8.8
USA golden retriever	Allelic	RS23932942	35	15274395	G	0.6	0.3	A	4.1	2.5	0.0425	1.6158	1.0	6.3
USA golden retriever	Allelic	RS9132002	29	22197558	G	0.2	0.0	A	3.9	6.9	0.0479	1.8198	0.8	60.1
USA GSD	Allelic	RS22184220	12	13111985	A	0.8	0.3	G	11.0	9.5	0.0009	0.0341	2.3	38.9
USA GSD	Allelic	RS23684720	31	35018207	T	0.5	0.1	G	7.9	7.7	0.0049	0.1845	1.7	35.0
USA GSD	Allelic	RS22859255	2	48522803	T	0.4	0.1	C	6.4	7.6	0.0117	0.4431	1.4	42.9
USA GSD	Allelic	RS24318716	6	4224712	C	0.1	0.4	T	5.4	0.1	0.0203	0.7703	0.0	0.9
USA GSD	Allelic	RS24872415	3	62441438	G	0.2	0.5	A	5.1	0.2	0.0238	0.9055	0.0	0.9
USA GSD	Allelic	RS22913552	20	33818500	G	0.7	0.3	T	4.4	3.8	0.0364	1.3821	1.1	13.5
USA GSD	Allelic	RS23690306	31	25600053	C	0.6	0.3	G	4.1	3.8	0.0436	1.6568	1.0	13.9
USA GSD	Allelic	RS22720467	19	32273778	C	0.0	0.2	T	3.9	0.0	0.0481	1.8289	0.0	NA
USA Labrador	Allelic	RS23026134	22	55676006	T	0.0	0.3	C	13.4	0.1	0.0003	0.0098	0.0	0.4
USA Labrador	geno/trend	RS23026134	22		-	-	-	-	-	-	0.0007	0.0253	-	-
USA Labrador	Allelic	RS24354997	6	14408543	T	0.3	0.1	G	6.7	3.3	0.0096	0.3653	1.3	8.3
USA Labrador	Allelic	RS24275059	6	14365278	C	0.3	0.2	A	6.0	2.7	0.0146	0.5563	1.2	6.1
USA Labrador	Allelic	RS22738169	2	45023367	T	0.6	0.4	C	5.9	2.3	0.0148	0.5632	1.2	4.5
USA Labrador	Allelic	RS23313105	26	23047328	C	0.4	0.2	T	5.1	2.4	0.0236	0.8960	1.1	5.0
USA Labrador	Allelic	RS23968934	36	32710839	T	0.3	0.5	C	4.9	0.5	0.0275	1.0446	0.2	0.9
USA Labrador	Allelic	RS22913552	20	33818500	G	0.4	0.6	T	4.3	0.5	0.0381	1.4467	0.3	1.0
USA Labrador	Allelic	RS23708599	31	30830928	A	0.5	0.4	G	4.1	2.0	0.0431	1.6370	1.0	3.9
USA Labrador	Allelic	RS9132002	29	22197558	G	0.0	0.0	A	4.0	NA	0.0468	1.7765	NA	NA
USA Labrador	Allelic	RS23545080	3	69975205	C	0.1	0.2	A	3.9	0.4	0.0469	1.7826	0.1	1.0

Appendix 11 continued

USA WHWT	Allelic	RS22915894	20	5771454	C	1.0	0.2	T	14.3	NA	0.0002	0.0060	NA	NA
USA WHWT	Allelic	RS8806978	20	5779740	C	1.0	0.2	T	14.3	NA	0.0002	0.0060	NA	NA
USA WHWT	geno/trend	RS22915894	20		-	-	-	-	-	-	0.0012	0.0443	-	-
USA WHWT	geno/trend	RS8806978	20		-	-	-	-	-	-	0.0012	0.0443	-	-
USA WHWT	Allelic	RS23932942	35	15274395	A	0.9	0.3	G	7.3	15.0	0.0069	0.2608	1.5	146.5
USA WHWT	Allelic	RS24556501	9	44666435	C	0.5	0.1	T	6.1	10.0	0.0132	0.5031	1.3	74.5
USA WHWT	Allelic	RS24408651	7	7829497	A	0.0	0.5	G	5.5	0.0	0.0195	0.7418	0.0	NA
USA WHWT	Allelic	RS23828846	33	29024878	T	0.1	0.6	C	5.1	0.1	0.0237	0.9006	0.0	0.9
USA WHWT	Allelic	RS24482628	8	27285603	C	0.1	0.5	A	4.2	0.1	0.0399	1.5151	0.0	1.1

Appendix 12 Candidate genes: Sequenom Primers and Probes

SNP	Reverse Primer	Forward Primer	Amplicon Length	Probe Sequence
rs8488679	ACGTTGGATGTGGCTG CTGGAAGCTTATTG	ACGTTGGATGACTGAG GGATGCAAAGCCTG	110	CTCCCTCCCGCTCCC
rs23828846	ACGTTGGATGACTGCA GACTTCCTATCTGG	ACGTTGGATGAGGTCA ATCAGACAGTGGTG	115	TCTGCCCCTCGACA A
rs22819627	ACGTTGGATGTTCCCA TCTGTGCCACAAAC	ACGTTGGATGCGATGA TTTCAGGAAAGGAG	101	ACAAACGGCTGCAT T
rs23602938	ACGTTGGATGAGAGG GTCCTGTTGATACAC	ACGTTGGATGGCCAGT AAGCTAGAAGTTCG	101	CCCATACTCCTCAG GC
rs22349092	ACGTTGGATGGGCTTT AGTAAGCAAGTCTG	ACGTTGGATGTTGTAG CTAAGTCCGCTCTG	100	AAGGACAGTGCGTT CA
rs22349093	ACGTTGGATGTCCGAC ACAAGAATGCAAAG	ACGTTGGATGCAGGG AAAATGACTCGAGAT	113	ATAACCCGTTGCAT CCC
rs22959786	ACGTTGGATGGGATGC AGACCGTCTGGATT	ACGTTGGATGGTCAGC AGACAACACTTAAC	115	CCGTCTGGATTTCCT TA
rs22509070	ACGTTGGATGGCCACG AGCAGTCAGGAGA	ACGTTGGATGTGGCCT GCTGTCCTTGATGT	80	TGGACAGAGCCAGA CAC
rs9070190	ACGTTGGATGCGCAG GATGTCCAGGTAGTT	ACGTTGGATGAGAGA CTTCTTCCCTTCCTG	117	GTCCAGGTAGTTCT GGTT
rs22000474	ACGTTGGATGTCTGCT TAACAGTCTCCTCC	ACGTTGGATGAGGCAT ATTAGATTGGCAGG	105	TGGCCACACCCATT CCTTC
rs24613432	ACGTTGGATGTAGGTC AAACCACTGATCCC	ACGTTGGATGGCCATC TGTCGTAACATTGC	102	CCGTGCTTGCTTTTT TTCC
rs24267550	ACGTTGGATGTCAATT TGAGCCCGGACTTG	ACGTTGGATGTCAAGA ATGGACCCTCTGAC	103	CGGGTAGATTCTCA TAGTC
rs22871627	ACGTTGGATGAGTCAG GTAGAGACTTACAG	ACGTTGGATGCACTTG CAGATGGAAGCAAC	110	CTCACATCTGGAGA GAGTT
rs23026134	ACGTTGGATGAGATCC TGGTTTGGAAGGAG	ACGTTGGATGCTTGGT GAGATCTCCTATGG	87	GGAGCATTGAGTGA GTCTAT
rs9150577	ACGTTGGATGAGATCT GTGAGCTCCTTGAG	ACGTTGGATGATGACA TCTATGTCTTCGGG	114	CTCGCGGTCACCCA CCGAGTT
rs24408651	ACGTTGGATGTCCTCT ACCAAATCTATGTG	ACGTTGGATGGCCAAC AGTATGCTGGTTAG	98	CAAATCTATGTGAA TAGCCCT
S100A8_p3	ACGTTGGATGACCTCC CTTAGCTTTCTCTC	ACGTTGGATGTCAGTC TGCTCCATGCTTTG	106	TCCCTTAGCTTTCTC TCCCCTCT
rs22986082	ACGTTGGATGCCCAGA ATAGTGAGAAACAG	ACGTTGGATGCTGCCT TGTGACTCTACTTG	101	GAATAGTGAGAAAC AGGATCAT
rs22328353	ACGTTGGATGACCTGC TACCAGCATTAAAG	ACGTTGGATGTTGCTT TCCTTCTCCCTTAG	99	GCTACCAGCATTAA AGCTTTAAC
rs8877521	ACGTTGGATGAATTCC AGCTCCCTCTTCAG	ACGTTGGATGGTGAA AACCCCTTACTGGAG	102	CTCCCTCTTCAGAA ACATCAGCCA
rs8806978	ACGTTGGATGGATGTG ACAAGCCTTTCTGC	ACGTTGGATGAGTCCG TGCTTTATCAAGCC	100	AGAGGCACCACATG TGCCCAGTTA

Appendix 12 continued

rs22931896	ACGTTGGATGTCCACT TGGAGAATACTGGG	ACGTTGGATGTGTAAT AAAGATGCCCCCAC	102	GGGAATTACTGCCA TCTGGTATAT
rs22605727	ACGTTGGATGTCCGTA GGAGGTTCTCTGG	ACGTTGGATGTGAGA ATGGAATCCTGGCAC	116	GGTTCTCTGGGCCC ACTGCCACTTC
rs22185491	ACGTTGGATGGGGAA TACAGCGGTCTTGA	ACGTTGGATGTTGTTC TCCTTCATCCTGGG	110	GGAATACAGCGGTC TTGACGGGTTG
rs23907319	ACGTTGGATGAGAGCT CAGTTTCCCATTGC	ACGTTGGATGTGAGA GGCACTGGAGATAAC	107	AAAGTAAGAAAGTT CTACAGGTAGG
rs8860216	ACGTTGGATGGTAAA ATCATGCACTGTCCC	ACGTTGGATGGTTCCT TGATTGTATTTACC	99	AAAATCATGCACTG TCCCACACAGTAG
rs23428471	ACGTTGGATGCCACAT GCTGATTACAAGCC	ACGTTGGATGGACCAC TTTTGTCTAGTAAG	117	CTATTGGAGTTGTAT TTATCGACAAAC
rs24657063	ACGTTGGATGACCACA GGTTAACCAGTGAC	ACGTTGGATGGCAGTG CGACTTTGTTGAAG	90	AACCAGTGACGAGA CTGATAGCCAATG
rs22511526	ACGTTGGATGTGGATT CCGAGTCTCGGTG	ACGTTGGATGATCAAG GACAGCAGGCCAC	80	ATTCCGAGTCTCGG TGGGCAGAGTCAGG
rs24328835	ACGTTGGATGCGATTG TGCCCATGTCCAAG	ACGTTGGATGGTAGAT CTCATACTGCTCCG	91	GCCCGCCAAGCCCT C
rs8555045	ACGTTGGATGTGTCCT TGGCGTGACCCTG	ACGTTGGATGATCCAG CACTAGGGAAAGAC	101	ACCCTGCGTGTCTTG
rs22542262	ACGTTGGATGACCAG ATCCATGTCCAGTTC	ACGTTGGATGATGGA GACTCAGATCTCAGC	118	GACTGCTGATGGTG G
rs22976618	ACGTTGGATGATCACC CCATCCCCTTCAC	ACGTTGGATGCTAGAC CCTTTGATGTGAAC	99	CTTCACCTCCAGGA AG
rs24332727	ACGTTGGATGTGTGGA GAGATGGATGGAAC	ACGTTGGATGATCCTT GAGCAACAGCCTTC	111	GTGAGGCTGTGTGT CC
rs22000453	ACGTTGGATGTTCCTC ACTGCTTAGGAGAC	ACGTTGGATGAAAGC GATTCCTAAATGCCC	99	AGGTCAAGGATGGT GG
rs24630609	ACGTTGGATGTGGACG ATGAGAGCCTCTTC	ACGTTGGATGACGTGG TCCCCTTGAATAAG	90	TCTTCCGGGCCGCA ATC
rs24404800	ACGTTGGATGGTTCAT TACAGCCACTGGAG	ACGTTGGATGAGCTGT GGTTTTGGAGCTGA	96	AACTTGGAGATCTT GGG
rs22655278	ACGTTGGATGAAATCC ATTTCCCCCTCTGC	ACGTTGGATGTCGTTC TTCCTCAGCTTCTC	106	AGACCTCCTCCTTA CTAC
rs23472497	ACGTTGGATGGGCCA GAGGATGGGAAATTC	ACGTTGGATGATCTCT GTCAGTGCACTAGG	99	ACCTGTCAAGGTCA TGTC
rs22212359	ACGTTGGATGAATCCC TCTTTTCCCCTCTG	ACGTTGGATGGCACA AGATGGAAGATGGAG	119	CCCCTCTGAAATTC CCTTT
rs24639522	ACGTTGGATGCCTGGA GGCAAAAGAAGATG	ACGTTGGATGGTAAGC ACTGATCCACTGAC	111	TCTGGAATGGAACA CGTCT
rs22542234	ACGTTGGATGATGACC AGATCCCTGTCTTC	ACGTTGGATGGCACA GAACCAATCATGGAC	102	CTGTCTTCTCCCAAT ACCAG
rs22655274	ACGTTGGATGCTAGAC AGCAGTTCAGTCTC	ACGTTGGATGTGGAG AAAGCCAGGCTTGCG	94	TTCAGTCTCCAGTCA GTGGG
rs24194054	ACGTTGGATGCAAAC AGATTGCTCCGTGTG	ACGTTGGATGGGTACA GAGACCAAGGAAAG	100	GTTGATGGAGAACA GCATGT

Appendix 12 continued

rs23066224	ACGTTGGATGGGGCTATTAGTGACCCAATG	ACGTTGGATGTCTGCAGTCAGATCAACTGT	94	CAGAAGTAGGATCATCTGCTC
rs22009467	ACGTTGGATGAGAGACCTCACCTCTGCTCC	ACGTTGGATGTCTGATCGTGGCATTCTTCC	92	TGCTCCCGTTCTCTCCCGTGTC
rs8649732	ACGTTGGATGGCTATTCAACTTTGGTGGCG	ACGTTGGATGTGGTTGTTGTCACTAGTTGG	101	CCTCACAAATCGGCTTAACTCT
rs22114085	ACGTTGGATGAAGCTCCACGATGAACAGTC	ACGTTGGATGGACTGTACACGACCTTTACC	115	CAGCAGCACTTCCAGAATGTACC
rs22615134	ACGTTGGATGAACCTTGGAACAAAGTCCCC	ACGTTGGATGTCAAGCTGGGTAGGTTAGTG	113	TTCAGCCAGAGTGTCCAGCCGTT
rs22588226	ACGTTGGATGTGGATCCCGAGTCTCGGTG	ACGTTGGATGAGGGTCACGCCAAGGACAG	88	CCGTATCGGTGGGCAGAGTCAGG
rs22212677	ACGTTGGATGTAACTTCTGCTCAGAAATG	ACGTTGGATGTGTGGTTTTGTATTCAAGGG	99	AATTGAGAATAATCAGTGAAATAG
rs22374487	ACGTTGGATGAACTCACACATCCCAGAGAG	ACGTTGGATGTCCAGTAATGACAGTGACAG	115	CCAGAGAGATCATTTTTATACTTTT
rs9255897	ACGTTGGATGTCTCCATGCGTTTCATGTCC	ACGTTGGATGTTTGTCCGGATGCAGCTATG	93	TTCATGTCCATCAGCTTCTCCACGTT
rs22599138	ACGTTGGATGGATCCGGTAACAGCAAACAC	ACGTTGGATGCTGAGTCTCTGAGTCTAGTG	103	CCGGGAATCCAGTGTCAGTCAGGCCA
rs23768387	ACGTTGGATGTCTCACTTCTCAGTCAGTGC	ACGTTGGATGGAAGAAAGAGATGAGGAGAG	99	TCACTTCTCAGTCAGTGCTGTCTCATA
rs9019919	ACGTTGGATGCACTGAGGACCAATTAGGTG	ACGTTGGATGTCCCACTATGAGCGAAACTG	101	CTGAGGACCAATTAGGTGTTTGAAATT
rs9095590	ACGTTGGATGTTCTTTCCTCTTCCCTCCTC	ACGTTGGATGGCAGAGCTCTTACCTACAAC	103	TTCCCTCCTCAGAATCATCTTCTCGAAC
rs22871630	ACGTTGGATGGGCAATAGATTAAGCCCAAG	ACGTTGGATGCTGCAAAGTTATATGTCTGG	119	ACTCCATACCTCTATGTAAAAACTGAAT
rs8857124	ACGTTGGATGGTGAACAGGACGTCCTCTC	ACGTTGGATGGGGGCCTGCAGATCAGTAT	87	CACAGCCACCGCCTC
rs24613434	ACGTTGGATGAGTCACTTGGGTATGTGTCG	ACGTTGGATGGGCTTGGATGCTGAGAAATG	108	TTTCACTCTGGCCCG
rs8555036	ACGTTGGATGAGTGGACGATTGCTGGTGG	ACGTTGGATGTGACAGCTCTAGGCACTCAG	111	CTGGTGGCGAGATCC
rs22542263	ACGTTGGATGCTGTCCTGTGACTGCTGATG	ACGTTGGATGTGACTCTGCCCATGGAGACT	95	TGCCCTTGTCTGTCTA
rs23545080	ACGTTGGATGCTTCTAACCTGAGTGAAGGC	ACGTTGGATGTGTGTTTCTGGAGTCCTACC	100	TGTGTTTGTGGGGTTA
rs24319791	ACGTTGGATGTGGCCTCGCAGATCCAGAAG	ACGTTGGATGAAGCAGCTGTCACCCAAAGG	116	ATCCAGAAGGCGCCGGA
rs22839343	ACGTTGGATGTATACTTCTCCATCAGCACC	ACGTTGGATGAACTTTCGGCAGTACGTGAG	103	TCCATCAGCACCACTCAC
rs24872415	ACGTTGGATGTTGGCAAGCCTGACAAATGG	ACGTTGGATGTGTTGAGGAGACTTCACCTG	100	ACCCAAGTGCCCAAACAA
rs24327271	ACGTTGGATGAGAACTCACTCTCTGTGGAC	ACGTTGGATGTGCCAAGAACAAGCCCAAAC	113	TCTGTGGACTAACTGAAA

Appendix 12 continued

rs21902631	ACGTTGGATGGTCTGT CTCTCTCCTTCTAC	ACGTTGGATGTATCTT ATGTTGAACATTTG	106	ATCTCCCAGTGATA ATTCC
rs22818977	ACGTTGGATGGATGGC AGGATGCATGGCAA	ACGTTGGATGAAGCTC CACTTTACCCATCC	120	GGATGCATGGCAAT AAATG
rs22874976	ACGTTGGATGCCACAT GCCTCTGACTTTTG	ACGTTGGATGAGACA ATCTGAGAACACAGC	85	CTCTGACTTTTGAGG AGTTC
rs23360288	ACGTTGGATGCAGCGT CAGACAGTTCTCCT	ACGTTGGATGAACCCA GCGGTGCTCACGTA	104	TCAGACAGTTCTCCT TCTCCA
rs24482628	ACGTTGGATGCAGGAT GCAGATTTTTCCAG	ACGTTGGATGACTTGG TGATTCTATGGAGG	96	TTCCAGTTTTCCCAT AGCCTG
rs22871624	ACGTTGGATGGCACTG GATTGAGACACTTC	ACGTTGGATGCTGGAG AGAGTTACTTAGAG	104	GGAAGCAACTAGAA TTTTAGC
rs22511518	ACGTTGGATGCACTGG ATTCCCGGTGTTTG	ACGTTGGATGCATCAG CAGTCACAGGACAG	99	ATTCCCGGTGTTTGC TGTTACC
rs22511504	ACGTTGGATGAGTCAG GATGCCCACGGGT	ACGTTGGATGAGACA GAGCCAGACACGCAG	82	AACGTGTGGCCTGC TGTCCTTG
rs24872954	ACGTTGGATGAAAGC CATCCATCCTCTCTG	ACGTTGGATGTCTGAG AGGGATCATGAGTG	117	TCCATCCTCTCTGTG CCACCTGT
rs22184220	ACGTTGGATGAGACC ATAGCCAGCTTGCAC	ACGTTGGATGTGGCAC AATCCCTTTGATGG	101	GAGTTAAAGGTATC TTCCATAGG
rs23122280	ACGTTGGATGGATTTC ACTCAGCCCCATAG	ACGTTGGATGAAACA GATGTCTCTAGTTGG	98	AGCCCCATAGAGCT TTCTTTAAAT
rs24438947	ACGTTGGATGTTTTCC TACTTCCCCACCTC	ACGTTGGATGTCTGGT TTCTCAACCAGGTG	99	CACCTCCCATCAAA ATTTTTCATTC
rs22908459	ACGTTGGATGCTTGAG GAATGTCCACCATC	ACGTTGGATGAGGAT GAAGCTTCTCGTAGG	101	ATGTCCACCATCTCT GGCTGCTGAC
rs22986033	ACGTTGGATGATAGGT AACACAAATTCTG	ACGTTGGATGATATTG ATCACGGAGCAGAC	92	GGTAACACAAATTC TGATTGCCTTG
rs24435410	ACGTTGGATGGACCAT TGGCCATTCTCTTG	ACGTTGGATGACCTGT AGCTTCCATTGTCC	100	TCTTGGCATGCTGTT CTCGCTGATCC
rs22989468	ACGTTGGATGAGGGA GTCACCTGCAAACTG	ACGTTGGATGATGCTC ACCTTGCGGAACTC	100	GCAAATGCAGTTGT GATGCAGTCGTG
rs24150067	ACGTTGGATGTTCGGC ACACAAAGTTCCTC	ACGTTGGATGATGGA ATGTATGCCCGCATC	95	GCACACAAAGTTCC TCACAAAGACATA
rs23274229	ACGTTGGATGGACTCA GGACCCACAATAAC	ACGTTGGATGACCTCC ATGGATCACTTCAG	119	AGGACCCACAATAA CAAAGGTCAAAAT
rs22750909	ACGTTGGATGAAGGA GAAGCAAATTCCCCC	ACGTTGGATGTTAGAA TTTGGTGCCCCCTC	116	CAAATTCCCCCCAC TCCAAAAAAAAAAA
rs21902676	ACGTTGGATGTTGGGC ATCTGTGTGGTTGG	ACGTTGGATGTCAGAC AGCTCCCATTTAGG	111	AGAGAGTCTCTGGT ACCCAGAGAACCCT
rs22655272	ACGTTGGATGAATTGT AGCCAACGTGCACC	ACGTTGGATGCCCATC TGGTTGTTCTAAGG	119	TGCACCGAGCACTG A
rs8867188	ACGTTGGATGTTTATC CTTGTCAGGTGCGG	ACGTTGGATGTTTAAC GTGGAGTGGGATGG	82	CAGGTGCGGTGCTT T
rs22864357	ACGTTGGATGAGATCT CGCACGTGTACTGG	ACGTTGGATGACTGAG CGACACCACTTCAC	111	CGTCCTCCACTACC AC

rs22523565	ACGTTGGATGAGACTG GATCTCTTGCTGTG	ACGTTGGATGGAAAA GGAGCTGTGAGGATG	100	CTTCACAAACTCTTG CC
rs24404799	ACGTTGGATGGTTCAT TACAGCCACTGGAG	ACGTTGGATGAGCTGT GGTTTTGGAGCTGA	96	ACTTGGAGATCTTG GGC
rs22859255	ACGTTGGATGGCATCT CAAATATTTCTGGG	ACGTTGGATGTCAGAG AAAGCCAAAAGCCG	90	TCTGGGTAAGAGGA ACA
rs21941198	ACGTTGGATGCCTGCC CATTGCTCATATTC	ACGTTGGATGGCTGCA GCTTGAAGATCAAC	97	GCTCGTCCACTCTAT TAG
rs22146864	ACGTTGGATGAATATG GTTGCAGGGCCTTC	ACGTTGGATGCTCTCC CTACTGATTTCCTC	101	GCCTTCCTCATACA CTCAC
rs23586992	ACGTTGGATGCACAGC CGTTGCTGTCCTT	ACGTTGGATGAGCACT GAGTTCCCACAGGC	120	CGTTGCTGTCCTTGG AGAA
rs22374481	ACGTTGGATGTTAGTC CCTCCTTGAGGTAG	ACGTTGGATGGGCAG GGCTCAACAGAAAAG	95	TGAGGTAGGAAAAC TCACA
rs24380950	ACGTTGGATGACTTTC TGGTACCCGTTTTG	ACGTTGGATGTATCTT CAATGAATGATGC	117	TCTTTTTTCCATCAG CAACT
rs22410121	ACGTTGGATGTTCTTC TCTGTCCTGTGACC	ACGTTGGATGTCTCAC AGCTGCTGTGAGCC	111	GCCGCTCCAGATCT TTGAACT
rs22509090	ACGTTGGATGTGCTGT CCTTGGCGTGACC	ACGTTGGATGAAGCC GCCACGAGCAGTCA	80	CCCGAGTGACCCTG CGTGTCT
rs8944179	ACGTTGGATGGAATG ACCAAGTAACTCTGC	ACGTTGGATGTATTCA TCAAGGACGCCAGC	94	AGTAACTCTGCTAA AATATGG
rs9013510	ACGTTGGATGGCGAG AAACACAGATTCTCT	ACGTTGGATGGAGAC AGACAGGCTATTTCC	117	TTTTTTTTCCTTTCC AATCTCA
rs22511502	ACGTTGGATGTGGCGG CTTCCTTGTCTTTC	ACGTTGGATGATCCTG ACTCTGCCCACCGA	81	CTTGTCTTTCCCTAG TGCTGGAT
rs9190946	ACGTTGGATGTCTTCT GTCTCCATCAGCAC	ACGTTGGATGGGAAA GCCTTTCTGCAGCGT	105	ATTGATGTAAAAAC CATGATGAA
S100A8_p2	ACGTTGGATGCCTGCA TATCCTCTGTCAAC	ACGTTGGATGCAAAG CATGGAGCAGACTGA	93	ATTTCGGGAAACCT GGTAAGTAA
rs22263793	ACGTTGGATGTAACTT ACCTTAGTATCTG	ACGTTGGATGCACCAG CTTAAAATCTACTG	100	ACTTACCTTAGTATC TGATATTTC
S100A8_P	ACGTTGGATGCAGTGA TTGCCCCATTCACC	ACGTTGGATGCCTAGG CCATGCAGAGATAG	103	TTGATTGCCCCATTC ACCTGGTTGA
rs23567144	ACGTTGGATGGATGA GCCCACACACAGAG	ACGTTGGATGTATCCC TGAAGCATCTGGGC	120	CCCACACACAGAGG CTGCTGGAATA
rs22986026	ACGTTGGATGTCCGGG TCACTCTTGGTAGA	ACGTTGGATGTCGGCC TAATGTTTAGGAGC	95	AGGCGGGAGCTGCG TGTCCACACAG
rs22871646	ACGTTGGATGGACAG GATACTGAAGGATGG	ACGTTGGATGGTTGTT CTGGGTCATGCTAC	87	TGAAGGATGGAAAC TAAATTGATGT
rs8667691	ACGTTGGATGATCCAA GCCCCTGGAGGTG	ACGTTGGATGTGTTGG AGATGCGCAGCG	100	ATGCTTCTGCTGGA GAAGAACCGCTC
rs22565034	ACGTTGGATGTGCACT ACTGCCACGTTGTC	ACGTTGGATGCACTAG ACTCAGAGACTCAG	100	TTCTCCTCTGACTGG ACCTGGACCGTT
rs22529256	ACGTTGGATGAAGAG TCTGGATCTGGAGAG	ACGTTGGATGTGACCA TCTGAGTTGGATTC	120	CGATTCTGGCTCCA ATAATTCTTCTAG

Appendix 12 continued

rs8988459	ACGTTGGATGGTCAGAATTACCAAGCTACG	ACGTTGGATGTCACTAGGTCTTACTTACCG	93	CAGAATTACCAAGCTACGGAAACAAGT
rs22915894	ACGTTGGATGGCCACTATCGTGTTGTCTAC	ACGTTGGATGCTGAGAATGGCTTCACAGAG	93	TGTCTACAGAACCAGATTTTGCAGAAGG
rs22807086	ACGTTGGATGAATGATGTAGCAGACAGGAC	ACGTTGGATGAGAGGGAACTGAACCCATGC	113	ATGATGGAATAAATCATTAGAACAAAAG
rs22655276	ACGTTGGATGTCTTCCTCCTAATCTCCTCC	ACGTTGGATGTGATGACCACCCTAGGACAC	99	CCACCCCCAAGCCAG
rs22542266	ACGTTGGATGACGGTGCTGAGATCTGAGTC	ACGTTGGATGACACACAGGGTCACAGCAAG	101	CAGGCTGCCAACGGG
rs22910701	ACGTTGGATGGACAAATACTTCCATGCCCG	ACGTTGGATGTGATCACTTTAGCAGCCCAG	99	CCGGGGGAACTATGA
rs24630571	ACGTTGGATGAGGAGGAAGAGGCCTCTGG	ACGTTGGATGGTGAAGCCACCAAGAGAGAC	114	AAGGCGGCGGTGGCGG
MS4A2_E5	ACGTTGGATGTATGCTGCTGACAGTGTTGG	ACGTTGGATGCTATTGAACGCACTGGCAGG	116	TCTCCCAGGCTGCCCCA
rs22029672	ACGTTGGATGATGCAGGGCTGGGTTTCTAC	ACGTTGGATGAGCAAGAGGAGCTATTCCTG	113	GGGTTTCTACCACGAAG
rs23690306	ACGTTGGATGTACTTCACCTGCAAGCTTGG	ACGTTGGATGGGTAGTTGAGGTAGATGAGG	103	AATAAGCCAATCCTCCTG
SPRR1B_E3_1	ACGTTGGATGAGGAAGTTTGGTGTTCCCAG	ACGTTGGATGACCTCAGGAGCCATTTGTTC	99	TTGCGGTAACTTGGTGTC
rs22509074	ACGTTGGATGCCAGGACAAGCAGACATCAG	ACGTTGGATGACGGGTCGTTAAGAATTCTC	100	GGGAGGAATCCAGCCTGA
rs21902632	ACGTTGGATGGAATGGTGGATGCCTTCAAC	ACGTTGGATGTAGACACCACAAGATCCTGC	95	ACGCCTTCAACACAAAGAG
rs9241373	ACGTTGGATGATGGGCCGCTGTCCTGCGT	ACGTTGGATGTCACGGAATTCTCCAAGGAC	115	CACTGCTCCGGCCCCCACAG
rs23593765	ACGTTGGATGCTGCAAGTATCTAAGGCCAC	ACGTTGGATGGGTATTCCAGGACTCTCAAG	99	CAAGGAGAGTTGGAGAGAG
rs24732893	ACGTTGGATGTTTCTTTGATTCTCTGGCGG	ACGTTGGATGTTGGTACCAGATCCAGCTAC	117	GGCCGGCGAAGATCCCCCAA
rs22509072	ACGTTGGATGAGTCAGGATGCCCACGGGT	ACGTTGGATGAGACAGAGCCAGACACGCAG	82	GTGTGGCCTGCTGTCCTTGA
rs22872666	ACGTTGGATGCGACCTTATGTGAAGGATGG	ACGTTGGATGTTTCCATCAGAGCCAAGGAC	91	GGATGGAAGACTTGTATGCAC
rs23478400	ACGTTGGATGTCTGCCTAGAATCTTGTCAC	ACGTTGGATGTTGCCCCCAGACATTTTTTG	118	GTACTTACATATGCATTTCCCA
rs24318716	ACGTTGGATGGTTGAACAGGACACCTCATC	ACGTTGGATGACTGTTGGAGGAGGCAGGAG	89	GAATCGGGCCCATTTGTTTCTA
rs22155657	ACGTTGGATGATGTTGAAGATTGAATGAC	ACGTTGGATGTATCTAAACCACCAGTTTC	113	TGTATATAAAGATTTCTCCTCAA
rs22935368	ACGTTGGATGTTTTCTGCGTGGGTTTAGGG	ACGTTGGATGTTATGGCCCAGATCCTTCTC	101	GAGTGAGGGCGGGCACCTTAATA
rs22913552	ACGTTGGATGTGAAACAAGATCACAAAGCC	ACGTTGGATGTGTCAGCACTGTGTGTGTTC	96	CCAAAGCCAAACTGAACAATAATC

Appendix 12 continued

rs22821176	ACGTTGGATGTATACA GACTGAATTGTTG	ACGTTGGATGGGAGA AAGAAGGAAAATAT	105	AAATTTATTTACTGC TAGTGTTAA
rs9196801	ACGTTGGATGCCCGTA CACAGTTGTCATTG	ACGTTGGATGTTCCTT GGTTCTGTAGGTGC	105	AGCGCGGGGCATTT AAATGGATTT
rs22511501	ACGTTGGATGATCCTG ACTCTGCCCACCGA	ACGTTGGATGTGGCGG CTTCCTTGTCTTTC	81	ACTATGCCCACCGA GACTCGGAATC
rs24264853	ACGTTGGATGTACCCA AACCCAAGAGGATG	ACGTTGGATGCCTATG CATATGACTAGCCC	103	GTACCTACATGTGA GTAATTGCTAA
rs22931897	ACGTTGGATGTAAAG ATGCCCCCACTTAGG	ACGTTGGATGTCCACT TGGAGAATACTGGG	97	AACTATGCCCCCAC TTAGGGCGATAT
rs22020166	ACGTTGGATGGTCTTT CCTTCTTAACCTGC	ACGTTGGATGGGGAA TCAAGAAAGGCTTTAC	120	TATTAACCTGCATA CAAAATTGTTGCT
rs9081246	ACGTTGGATGGACCCA AAGATATTCTGACC	ACGTTGGATGCCGCCC ATGTCTCATGAATA	104	AAAGATATTCTGAC CTAAAATACTATG
rs9013509	ACGTTGGATGCAAATT TAGTAGAAGATAGAG	ACGTTGGATGTCTCCC AGTGATAATTCCTC	114	AGAAGATAGAGTTT TAAAATATTTTTTT
rs8555046	ACGTTGGATGTGGCGG CTTCCTTGTCTTTC	ACGTTGGATGAGCCTG ACTCTGCCCACCGA	81	TCCCTAGTGCTGGA T
rs22181912	ACGTTGGATGTTCTGC AAGGATGGGCATTC	ACGTTGGATGTCCAGT TTTTCTATGTCCTG	102	AGAACATGTCCTTG TGG
rs24202381	ACGTTGGATGACCGGC TCAATGATGATGTC	ACGTTGGATGAGCTTC AGCTCCTCTCTTAG	99	TGTCAAGCGCTATA GCTG
rs22588227	ACGTTGGATGAGTCAG GATGCCCACGGGT	ACGTTGGATGAGACA GAGCCAGACACGCAG	82	ACCCCGGGTGGCCT GCTGT
rs22217971	ACGTTGGATGATCCCG AGCCCCTGCCGGT	ACGTTGGATGGAAGTC CAGGTGGTGGAATG	115	GCCGGTGGTGCCTC C
rs22509067	ACGTTGGATGTGGATC CCGAGTCTCGGTG	ACGTTGGATGATCAAG GACAGCAGGCCAC	80	GTGGGCAGAGTCAG G
rs8555044	ACGTTGGATGTGCTGT CCTTGGCGTGACC	ACGTTGGATGAAGCC GCCACGAGCAGTCA	80	TTGACCCTGCGTGTC T
rs8877519	ACGTTGGATGCGGAG AGGTACTTCTGAAAG	ACGTTGGATGGAGAG AATTTTTCTCAGGGC	100	AGTTCAATTTCAGG TGC
rs22542260	ACGTTGGATGTGACTG ACACTGGATTCCCG	ACGTTGGATGCATCAG CAGTCACAGGACAG	106	TTCCCGGTGTTTGCT
rs22588228	ACGTTGGATGAGACA GAGCCAGACACGCAG	ACGTTGGATGAGTCAG GATGCCCACGGGT	82	CGACACGCAGGGTC AC
rs22908463	ACGTTGGATGGTCTCC TTTTCATGTGTCCC	ACGTTGGATGAGGAC CAGGGAGCACAATAG	98	GGGATGAAGCTTCT CGT

Appendix 13 Candidate gene study: Individual breed analysis (p<0.05)

Breed	Statistical Test	Marker	Chr	Bp	A1	F_A	F_U	A2	Chisq	Orx	P	Correct P Value (X48)	L95	U95
Shiba Inu	Allelic	Rs22986026	21	53629181	C	0.5	0.1	T	18.4	7.4	0.0000	0.0009	2.8	19.9
Usa Whwt	Allelic	Rs23274229	25	6184255	C	1	0.1	T	17.6	NA	0.0000	0.0010	NA	NA
Uk Labrador	Allelic	Rs23907319	36	10444284	C	0.4	0.1	T	19.4	6.4	0.0000	0.0005	2.6	15.4
Shiba Inu	Allelic	Rs9070190	21	28910870	G	0.8	0.3	A	24	9.2	0.0000	0.0001	3.6	23.4
Shiba Inu	Allelic	Rs9150577	21	28912392	A	0.2	0.7	G	21.3	0.1	0.0000	0.0002	0.1	0.3
Uk golden retriever	Trend	Rs22542260	17	-	-	-	-	-	15.1	-	0.0001	0.0050	-	-
Uk golden retriever	Allelic	Rs22542260	17	64296760	A	0.3	0	G	14.5	19.8	0.0001	0.0067	2.6	153.7
Uk Labrador	Trend	Rs22588227	17	-	-	-	-	-	16.1	-	0.0001	0.0030	-	-
Pit Bull	Allelic	Rs24613432	39	751050	A	0	0.4	G	15.1	0	0.0001	0.0048	0	0.3
Shiba Inu	Trend	Rs9070190	21	-	-	-	-	-	15.5	-	0.0001	0.0040	-	-
Shiba Inu	Trend	Rs9150577	21	-	-	-	-	-	14.4	-	0.0001	0.0070	-	-
Uk golden retriever	Trend	Rs22542263	17	-	-	-	-	-	13.8	-	0.0002	0.0100	-	-
Uk Labrador	Allelic	Rs22588227	17	64297022	G	0.3	0.1	C	13.8	5.6	0.0002	0.0098	2.1	14.9
Uk Labrador	Add	Rs22588227	17	64297022	-	-	-	-	-	8.7	0.0002	0.0099	2.8	27.1
Uk Labrador	Trend	Rs23907319	36	-	-	-	-	-	13.7	-	0.0002	0.0110	-	-
Uk golden retriever	Allelic	Rs22542263	17	64296837	T	0.3	0	C	13.2	18.2	0.0003	0.0135	2.3	140.9
Usa German Shepherd Dog	Allelic	Rs22807086	2	45212191	A	0.7	0.1	G	13.1	14	0.0003	0.0138	3	66.4
Shiba Inu	Allelic	Rs24328835	6	10890690	T	0.7	0.3	C	13.4	4.6	0.0003	0.0123	2	10.6
Uk golden retriever	Allelic	Rs24613434	39	750979	G	1	0.3	A	12.4	NA	0.0004	0.0201	NA	NA
Shiba Inu	Add	Rs9070190	21	28910870	-	-	-	-	-	4.7	0.0004	0.0181	2	11
Usa Whwt	Allelic	Rs22009467	10	23005117	A	0.5	0	G	11.7	Na	0.0006	0.0310	Na	Na
Uk golden retriever	Allelic	Rs22871630	2	45133611	A	0.5	0.1	C	11.7	12.1	0.0006	0.0296	2.1	68.9
Shiba Inu	Add	Rs9150577	21	28912392	-	-	-	-	-	0.2	0.0006	0.0264	0.1	0.5
Shiba Inu	Trend	Rs22986026	21	-	-	-	-	-	11.4	-	0.0007	0.0360	-	-
Pit Bull	Trend	Rs24613432	39	-	-	-	-	-	11.4	-	0.0008	0.0360	-	-

Appendix 13 continued

Pit Bull	Allelic	Rs24319791	6	10917564	A	0.3	0	G	11	NA	0.0009	0.0440	NA	NA
Uk golden retriever	Trend	Rs22871630	2	-	-	-	-	-	10.9	-	0.0010	0.0470	-	-
Uk Labrador	Add	Rs23907319	36	10444284	-	-	-	-	-	3.9	0.0010	0.0501	1.7	9
Usa Labrador	Allelic	RS21902632	1	16476102	G	0.1	0.3	C	10.4	0.2	0.0013	0.0605	0.1	0.6
Usa golden retriever	Allelic	RS22819627	2	45159024	T	0.6	0.3	A	9.9	4.3	0.0016	0.0775	1.7	11.0
Uk Labrador	Allelic	RS24657063	39	755447	G	0.7	0.4	A	9.5	3.1	0.0021	0.1007	1.5	6.5
Uk golden retriever	Allelic	RS8555046	17	64298495	T	0.3	0.0	C	9.5	11.7	0.0021	0.1011	1.7	79.7
Uk Labrador	Allelic	RS22818977	2	45167469	C	0.3	0.1	T	9.2	3.7	0.0024	0.1149	1.5	8.9
Uk Labrador	Allelic	RS22871630	2	45133611	A	0.3	0.1	C	8.9	4.0	0.0028	0.1340	1.5	10.6
Uk golden retriever	Allelic	RS22819627	2	45159024	T	0.5	0.1	A	8.9	9.2	0.0028	0.1356	1.7	50.6
Uk golden retriever	Allelic	RS22871627	2	45133465	T	0.5	0.1	C	8.9	9.2	0.0028	0.1356	1.7	50.6
Uk golden retriever	Allelic	RS22871646	2	45133786	G	0.5	0.1	A	8.9	9.2	0.0028	0.1356	1.7	50.6
Uk golden retriever	Allelic	RS22872666	2	45162084	T	0.5	0.1	G	8.9	9.2	0.0028	0.1356	1.7	50.6
Uk golden retriever	Allelic	RS22874976	2	45176838	A	0.5	0.1	G	8.9	9.2	0.0028	0.1356	1.7	50.6
Uk Labrador	Allelic	RS22374481	14	46159795	C	0.2	0.0	T	8.9	5.2	0.0029	0.1380	1.6	16.9
Shiba Inu	Allelic	RS21902676	1	16469324	G	0.3	0.6	A	8.8	0.3	0.0029	0.1409	0.1	0.7
Shiba Inu	Allelic	RS21941198	1	16475961	C	0.3	0.6	A	8.8	0.3	0.0029	0.1409	0.1	0.7
Uk golden retriever	Allelic	RS24657063	39	755447	G	1.0	0.4	A	8.7	NA	0.0032	0.1547	NA	NA
Usa Labrador	Allelic	RS9013510	1	16476441	A	0.1	0.0	G	8.6	12.8	0.0034	0.1631	1.5	107.7
Uk golden retriever	Allelic	RS22818977	2	45167469	C	0.5	0.1	T	8.5	8.8	0.0035	0.1687	1.6	48.7
Usa golden retriever	Allelic	RS24613432	39	751050	A	0.0	0.2	G	8.4	0.0	0.0038	0.1801	0.0	NA
Usa German Shepherd Dog	Allelic	MS4A2_E5	21	53634435	A	0.0	0.4	G	8.2	0.0	0.0042	0.2031	0.0	NA
Uk golden retriever	Allelic	RS22807086	2	45212191	A	0.5	0.1	G	8.1	8.4	0.0044	0.2132	1.5	45.9
Uk Labrador	Allelic	RS9013510	1	16476441	A	0.2	0.1	G	8.0	4.6	0.0047	0.2239	1.5	14.2
Uk Labrador	Allelic	RS22542263	17	64296837	T	0.1	0.0	C	7.0	5.7	0.0083	0.3962	1.4	24.1
Uk Labrador	Allelic	RS8555046	17	64298495	T	0.2	0.0	C	6.7	4.4	0.0095	0.4560	1.3	14.8
Shiba Inu	Allelic	RS22529256	17	64273415	T	0.1	0.0	C	6.7	NA	0.0096	0.4632	NA	NA

Appendix 13 continued

Usa German Shepherd Dog	Allelic	RS22655278	9	5853780	G	0.0	0.3	A	6.3	0.0	0.0121	0.5794	0.0	NA
Uk Labrador	Allelic	RS22871646	2	45133786	G	0.3	0.1	A	6.2	3.1	0.0126	0.6029	1.2	7.6
Japan golden retriever	Allelic	RS8877521	21	53631234	C	0.2	0.5	T	6.1	0.2	0.0138	0.6634	0.1	0.8
Uk Labrador	Allelic	RS24328835	6	10890690	C	0.3	0.1	T	6.1	2.8	0.0139	0.6658	1.2	6.4
Shiba Inu	Allelic	RS22872666	2	45162084	T	0.3	0.1	G	6.0	3.4	0.0144	0.6931	1.2	9.4
Usa Labrador	Allelic	RS24630609	39	745403	G	0.2	0.0	A	6.0	4.8	0.0146	0.7003	1.2	18.8
Uk German Shepherd Dog	Allelic	RS9019919	21	53636716	A	0.1	0.0	G	5.9	NA	0.0149	0.7128	NA	NA
Boxer	Allelic	RS8555046	17	64298495	T	0.0	0.3	C	5.9	0.0	0.0155	0.7459	0.0	NA
Usa golden retriever	Allelic	RS22986082	21	53634938	G	0.2	0.5	A	5.8	0.3	0.0161	0.7747	0.1	0.8
Uk golden retriever	Allelic	RS22588227	17	64297022	G	0.3	0.1	C	5.6	7.0	0.0184	0.8813	1.1	44.2
Shiba Inu	Allelic	RS22986082	21	53634938	G	0.2	0.5	A	5.5	0.4	0.0191	0.9158	0.2	0.9
Usa German Shepherd Dog	Allelic	RS23907319	36	10444284	C	0.4	0.1	T	5.4	5.6	0.0198	0.9504	1.2	25.8
Uk German Shepherd Dog	Allelic	RS21902632	1	16476102	G	0.1	0.4	C	5.3	0.1	0.0207	0.9955	0.0	0.9
Boxer	Allelic	RS23274229	25	6184255	T	0.2	0.0	C	5.3	NA	0.0216	1.0354	NA	NA
Uk Labrador	Allelic	SPRR_E3_1	17	64883377	T	0.1	0.3	C	5.2	0.3	0.0231	1.1088	0.1	0.9
Usa golden retriever	Allelic	RS22871627	2	45133465	T	0.4	0.2	C	4.9	3.2	0.0261	1.2547	1.1	9.4
Uk golden retriever	Allelic	RS22542262	17	64296821	T	0.3	0.1	C	4.9	6.3	0.0268	1.2854	1.0	39.1
Usa golden retriever	Allelic	RS22986026	21	53629181	C	0.5	0.3	T	4.7	2.8	0.0299	1.4352	1.1	7.2
Uk Labrador	Allelic	RS22819627	2	45159024	T	0.2	0.1	A	4.7	2.7	0.0306	1.4698	1.1	6.8
Uk Labrador	Allelic	RS8877519	21	53631150	G	0.5	0.4	A	4.7	2.1	0.0308	1.4774	1.1	4.3
Japan Shih Tzu	Allelic	RS24613432	39	751050	A	0.0	0.1	G	4.6	0.1	0.0318	1.5269	0.0	1.1
Uk Labrador	Allelic	MS4A2_E5	21	53634435	A	0.2	0.4	G	4.6	0.4	0.0323	1.5504	0.2	0.9
Pit Bull	Allelic	RS24613434	39	750979	G	0.4	0.6	A	4.6	0.4	0.0324	1.5542	0.1	0.9
Shiba Inu	Allelic	RS22931896	20	9213048	A	0.3	0.1	G	4.5	3.1	0.0342	1.6392	1.1	9.2
Japan Shih Tzu	Allelic	RS22146864	11	23976639	G	0.2	0.3	A	4.5	0.5	0.0345	1.6555	0.2	1.0
Usa Labrador	Allelic	RS21941198	1	16475961	C	0.2	0.4	A	4.4	0.4	0.0365	1.7496	0.2	1.0
Uk German Shepherd Dog	Allelic	RS22986082	21	53634938	G	0.6	0.3	A	4.3	3.7	0.0382	1.8341	1.0	13.5

Appendix 13 continued

Uk German Shepherd Dog	Allelic	RS21941198	1	16475961	C	0.1	0.5	A	4.3	0.2	0.0383	1.8379	0.0	1.0
Boxer	Allelic	RS21941198	1	16475961	A	0.1	0.4	C	4.3	0.2	0.0383	1.8403	0.0	1.0
Shih Tzu	Allelic	RS24613432	39	751050	A	0.0	0.1	G	4.3	0.1	0.0384	1.8446	0.0	1.2
Japan golden retriever	Allelic	RS24264853	5	71503539	G	0.5	0.2	A	4.2	3.3	0.0407	1.9526	1.0	10.4
Japan golden retriever	Allelic	RS8877519	21	53631150	A	0.2	0.5	G	4.2	0.3	0.0407	1.9526	0.1	1.0
Shiba Inu	Allelic	RS22871627	2	45133465	T	0.3	0.1	C	4.2	2.9	0.0408	1.9574	1.0	8.2
Boxer	Allelic	MS4A2_E5	21	53634435	A	0.4	0.1	G	4.2	4.5	0.0411	1.9728	1.0	20.3
Boxer	Allelic	RS8877519	21	53631150	A	0.4	0.1	G	4.1	4.0	0.0425	2.0410	1.0	16.0
Boxer	Allelic	RS8877521	21	53631234	C	0.4	0.1	T	4.1	4.0	0.0425	2.0410	1.0	16.0
Uk German Shepherd Dog	Allelic	RS22871646	2	45133786	G	0.2	0.5	A	4.1	0.2	0.0436	2.0947	0.1	1.0
Uk golden retriever	Allelic	RS9190946	3	15784064	T	0.8	0.4	C	4.1	7.0	0.0439	2.1082	0.8	62.2
Uk German Shepherd Dog	Allelic	RS22871630	2	45133611	A	0.2	0.5	C	4.1	0.2	0.0441	2.1144	0.1	1.0
Uk golden retriever	Allelic	MS4A2_E5	21	53634435	A	0.0	0.4	G	4.0	0.0	0.0453	2.1725	0.0	NA
Japan Shih Tzu	Allelic	RS21902632	1	16476102	G	0.1	0.1	C	4.0	0.3	0.0458	2.2003	0.1	1.0
Pit Bull	Allelic	RS9190946	3	15784064	T	0.2	0.1	C	4.0	4.7	0.0459	2.2018	0.9	23.4

Appendix 14 Candidate gene study: Individual Breed Meta-analysis (p<0.05)

Breed	SNP	Chromosome	Base Pairs	Odds Ratio	P Value	Corrected P Value (X48)	L95	U95
Labradors	RS21902632	1	16476102	0.4	0.006	0.26635	0.2	0.8
Whwt	RS23274229	25	6184255	5.4	0.008	0.36437	1.5	19.4
German Shepherd Dog	RS22655278	9	5853780	0.0	0.009	0.42302	NA	NA
golden retriever	RS22986082	21	53634938	0.4	0.012	0.57072	0.2	0.8
golden retriever	SPRR1B_E3_1	17	64883377	0.3	0.013	0.62064	0.1	0.8
German Shepherd Dog	RS21902632	1	16476102	0.2	0.013	0.63024	0.1	0.7
Labradors	RS21941198	1	16475961	0.5	0.018	0.84720	0.2	0.9
Labradors	RS21902676	1	16469324	0.5	0.026	1.22688	0.3	0.9
German Shepherd Dog	RS22818977	2	45167469	0.4	0.043	2.04384	0.1	1.0
German Shepherd Dog	RS8877519	21	53631150	0.3	0.045	2.16096	0.1	1.0
German Shepherd Dog	RS22807086	2	45212191	2.5	0.048	2.32608	0.9	6.4
German Shepherd Dog	RS22374481	14	46159795	3.7	0.051	2.42640	0.9	14.5

Publications

Wood SH, Ke X, Nuttall T, McEwan N, Ollier WE, Carter SD. (2009), Genome-wide association analysis of canine atopic dermatitis and identification of disease related SNPs. *Immunogenetics*, 61(11-12):765-72

Wood SH, Clements DN, Ollier WE, Nuttall T, McEwan NA, Carter SD. (2009), Gene expression in canine atopic dermatitis and correlation with clinical severity scores. *Journal of Dermatological Science*, 55(1):27-33

Wood SH, Clements DN, McEwan NA, Nuttall T, Carter SD. (2008), Reference genes for canine skin when using quantitative real-time PCR. *Veterinary Immunology and Immunopathology*, 126(3-4):392-5

Merryman-Simpson AE, **Wood SH**, Fretwell N, Jones PG, McLaren WM, McEwan NA, Clements DN, Carter SD, Ollier WE, Nuttall T. (2008), Gene (mRNA) expression in canine atopic dermatitis: microarray analysis. *Veterinary Dermatology*, 19(2):59-66.