Bioinformatic screening for candidate mutations underlying phenotypic traits in domestic animals

Shumaila Sayyab

Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics
Uppsala

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2014

Acta Universitatis agriculturae Sueciae
2014:92

ISSN 1652-6880
ISBN (print version) 978-91-576-8132-4
ISBN (electronic version) 978-91-576-8133-1

Print: SLU Service/Repro, Uppsala 2014

Bioinformatic screening for candidate mutations underlying phenotypic traits in domestic animals

Abstract

Domestic animals represent excellent model organisms for gene mapping and identification of mutations underlying phenotypic traits. Humans have selected spontaneous mutations in farm and companion animals since they were domesticated and this has resulted in large phenotypic variation among different breeds. In this thesis, we evaluate the candidate mutations in domesticated animals from NGS and SNP genotype data using bioinformatic analysis. Functional significance of coding sequence polymorphisms was assessed using both available bioinformatics resources and in-house pipelines. In consequence, pig and rabbit sequencing revealed major sweeps for genes (*NR6A1*, *LCORL* and *PLAG1*) for body length and increased number of vertebrae in domestic pigs and genes (*GRIK2* and *SOX2*) affecting brain and neuronal development in rabbit domestication. Genome-wide association mapping for demodicosis disease in Staffordshire Bull Terrier dog shows several preliminary candidate risk loci (CFA17, 18, 28 and 29) containing interesting candidate genes providing a good basis for further evaluation. Additionally, we also highlight some opportunities and pitfalls of whole genome re-sequencing using the Ion Proton platform and developed a tool DevRO (using deviant read paired orientation) for detection of large structural variants for NGS data from paired-end sequencing or mate pair. This method will be useful when large numbers of populations are re-sequenced as compared to traditional methods that can detect the structural variants in a pair-wise manner.

Keywords: Bioinformatics, Next generation sequencing, Genome-wide association, SNP, Structural Variations.

Author's address: Shumaila Sayyab, Department of Animal Breeding and Genetics,
SLU, P.O. Box 7023, 750 07 Uppsala, Sweden
E-mail: shumaila.sayyab@slu.se

Dedication

To my family and teachers, especially to my dear father and loving husband for their trust, motivation and support.

Contents

List of Publications

I Carl-Johan Rubin*, Hendrik-Jan Megens*, Alvaro Martinez Barrio, Khurram Maqbool, **Shumaila Sayyab**, Doreen Schwochow, Chao Wang, Örjan Carlborg, Patric Jern, Claus Jørgensen, Alan L. Archibald, Merete Fredholm, Martien A. M. Groenen, and Leif Andersson (2012). Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci USA* 109(48), 19529-36

II Miguel Carneiro*, Carl-Johan Rubin*, Federica Di Palma*, Frank W. Albert, Jessica Alföldi, Alvaro Martinez Barrio, Gerli Pielberg, Nima Rafati, **Shumaila Sayyab**, Jason Turner-Maier, Shady Younis, Sandra Afonso, Bronwen Aken, Joel M. Alves, Daniel Barrell, Gerard Bolet, Samuel Boucher, Hernán A. Burbano, Rita Campos, Jean L. Chang, Veronique Duranthon, Luca Fontanesi, Hervé Garreau, David Heiman, Jeremy Johnson, Rose G. Mage, Ze Peng, Guillaume Queney, Claire Rogel-Gaillard, Magali Ruffier, Steve Searle, Rafael Villafuerte, Anqi Xiong, Sarah Young, Karin Forsberg-Nilsson, Jeffrey M. Good, Eric S. Lander, Nuno Ferrand*, Kerstin Lindblad-Toh*, Leif Andersson* (2014). Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 345, 1074-1079

III **Shumaila Sayyab**, Susanne Åhman, Katarina Varjonen, Tomas Bergström, Kerstin Bergvall, Frode Lingaas, Göran Andersson (2014). Genome-wide association mapping identifies candidate risk loci for a canine juvenile-onset demodicosis disease in Staffordshire Bull Terriers. (Manuscript)

IV Agnese Viluma*, **Shumaila Sayyab***, Sofia Mikko, Göran Andersson, Tomas Bergström (2014). Whole genome sequencing for variant detection of four chinese crested dogs. (Submitted)

V **Shumaila Sayyab**, Nima Rafati, Miguel Carneiro, Göran Andersson, Leif Andersson, Carl Johan Rubin. Computational method for detection of structural variants using Deviant Reads and read pair Orientation: DevRO. (Manuscript)

Papers I-II are reproduced with the permission of the publisher.

The contribution of Shumaila Sayyab to the papers included in this thesis was as follows:

I Contributed in bioinformatic analysis of functional significance of candidate causative mutations in coding region and participation in writing.

II Contributed in bioinformatic analysis of coding SNPs and participation in writing.

III Involved in planning and performed GWAS analysis of demodicosis disease in Staffordshire Bull terriers and wrote the first draft of the manuscript.

IV Participated in planning of the study and performed analysis of whole genome sequences from dog trios to identify causative mutations for monogenic diseases. Contributed in writing of the manuscript.

V Involved in design of the study and developed the software DevRO for detection of structural variants. Wrote the first draft of the manuscript.

Related Work by Author

(Not included in the thesis)

- Shumaila Sayyab, Agnese Viluma, Kerstin Bergvall, Emma Brunberg, Tosso Leeb, Tomas Bergström, Göran Andersson. Whole Genome Sequencing of a Kromfohrländer Family Trio Reveals a Causative Mutation for Footpad Hyperkeratosis. (To be submitted)

- Khurram Maqbool, Shumaila Sayyab, Alvaro Martinez Barrio, Bertrand Bed'hom, Michele Tixier-Boichard, Paul Siegel, Carl-Johan Rubin and Leif Andersson. Detecting signatures of selection in broiler and layer chickens using whole genome resequencing of pooled samples. (Manuscript)

Abbreviations

AF	allele frequency
BLAST	basic local alignment search tool
BLAT	BLAST-like alignment tool
bp	base pair
CFA	Canis familiaris chromosome
CNV	copy number variation
DLA	dog leukocyte antigen
DoC	depth of coverage
GWAS	Genome-wide association study
INDEL	Insertion or deletion
kb	kilo bases
LD	linkage disequilibrium
MAF	minor allele frequency
Mb	mega bases
MHC	major histocompatibility complex
MP	mate pair
NGS	Next generation sequencing
PE	Paired-end
PEM	paired end mapping
QTL	quantitative trait locus
SNP	Single Nucleotide Polymorphism
SV	Structural Variations
UTR	untranslated region
WES	whole exome sequencing
WGS	whole genome sequencing

1 Introduction

1.1 Genetics through history

The story of genetics started during the years 1856-1863 when the father of genetics Gregor Mendel studied "trait inheritance". Since prehistoric times, Mendel's observation that living things inherit traits from their parents has been used to improve crop plants and animals through selective breeding (Weiling, 1991). The importance of Mendel's work did not gain wide understanding until the mid 19th century. After the rediscovery of Mendel's work, several discoveries and findings have been made that contributed to the history of modern genetics, including the structure of DNA established as a double helix in 1953 (Watson & Crick 1953).

The widespread use of markers to define genetic variation and development of new technologies made it possible to associate disease phenotypes with genetic loci. Monogenic diseases (*i.e.* when a mutation in a single gene is responsible for disease) provided an invaluable opportunity to learn about underlying molecular mechanisms. Single gene diseases run in families and can be sex-linked or autosomal dominant or recessive. The knowledge of pedigree analyses of large families and many affected individuals has been used to determine if the disease gene is sex-linked or autosomal and disease phenotype is recessive or dominant. Traditional linkage mapping, candidate gene approaches and Genome-Wide Association Studies (GWAS) have been used to identify mutations for Mendelian disorders. An alternative approach is to use whole-genome sequencing (WGS) with Next Generation Sequencing (NGS) technologies. This has been successfully explored in human diseases where single genes underlie rare Mendelian disease. A few examples of monogenic diseases in human that have an autosomal recessive mode of inheritance include Sickle-cell anemia, Cystic fibrosis and Albinism. Huntington's disease (autosomal dominant mode), while Hemophilia A and

Rett's syndrome are X-linked with recessive and dominant mode of inheritance, respectively. On the other hand, complex diseases do not follow standard Mendelian patterns of inheritance and are caused by a combination of multiple genetic and environmental factors. Alzheimer's disease, diabetes, cancer, heart disease, asthma, Parkinson's disease, multiple sclerosis, osteoporosis, and autoimmune diseases are examples of complex diseases in humans (Hunter, 2005).

Following extensive linkage mapping and physical mapping using polymorphic genetic markers and through genome sequencing using Sanger sequencing technology, a draft version of the human genome was made available in 2000 (Lander *et al.*, 2001; Venter *et al.*, 2001). With the completion of the Human Genome Project in 2003, researchers began to identify the areas in the genome that differ between individuals. Dramatic inroads have been made into the study of polygenic and complex human diseases. As part of these sequencing efforts, a large number of single nucleotide polymorphisms (SNPs) were identified and SNP chips were developed (Sachidanandam *et al.*, 2001). The International Hapmap project mapped these SNPs along the length of every human chromosome (International HapMap *et al.*, 2010; Sachidanandam *et al.*, 2001).

These technologies have allowed the researchers to efficiently map the disease to a gene and associate the phenotype to the genotype. Today, NGS technologies produce increasing amount of data that allows scientists to sequence the whole genome of mammals and many other species at a very low cost. This has become an attractive alternative to SNP chips where the disease is rare and only a few individuals are available.

1.2 Domestic animals as models

The phenotypic selection in farm and companion animals has resulted in a wide diversity of breeds of domestic animals since they were domesticated. In 1859, Charles Darwin was the first to recognize this phenotypic diversity in crops and domestic animals that occurred due to breeding (Darwin, 1859). This artificial selection for certain traits beneficial for man over several thousand years has led to many changes in domestic animals such as external morphology (coat colour, fur type, body size, smaller skulls and legs), internal morphology (brain size, smaller intestines), physiological changes (endocrine response, reproductive cycle), developmental changes (earlier sexual maturity), behavioral changes (decrease in fear and anti-predator response, increase in sociability) etc.

Domestic animals therefore represent excellent model organisms for identifying the genes that control growth, reproduction, behavior, development and appetite and several other traits for which they are bred (Andersson, 2001; Figure 1).

There are several reasons why domestic animals are good models (Andersson, 2009; Georges, 2007):

(i) Large pedigrees are easily accessible due to excellent record of pedigree information.
(ii) Easier to collect tissue samples.
(iii) Selection for desirable phenotypes
(iv) Limited genetic variation within breed
(v) Excellent record keeping of phenotypes
(vi) Similarities between diseases of domestic animals and human (*e.g.* dogs and human have many identical diseases like cancer, epilepsy and allergies etc.)

1.3 Accumulation of mutations

Humans have performed selective breeding for thousands of years, with an increase in intensity during the recent years (Andersson, 2001). This artificial selection over several generations has resulted in the accumulation of both advantageous and deleterious mutations. Some of these mutations have very obvious phenotypic effect (*e.g.* coat color) while others have subtle effect on traits like disease resistance, production, fertility and behavior.

The deleterious mutations are usually eliminated from the population due to purifying selection while advantageous mutations become more common in a population due to strong positive selection leading to favorable desired traits. Farm animals contain a wealth of mutations in genes that cause morphological, behavioral, reproductive and physiological changes. Research has been performed extensively on in particular reproduction and production traits in these animals (*Online Mendelian Inheritance in Animals (OMIA)*).

A few examples of causative mutations in domestic animals that resulted in changed phenotype due to selections for these traits include:

- ➢ *RYR1* gene (g.1843C>T) for lean meat and muscularity in pigs and 10 other species
- ➢ *PRKAG3* gene (R200Q) for increase in glycogen content in skeletal muscle in pigs

- *IGF2* (intron-3 G>A) for increase in muscle growth, heart size and decrease in fat deposition in pigs.
- *MSTN* (loss of function mutation in cattle) for double muscling.
- *DGAT1* (K232A) for effect on milk fat content in cattle.
- *BMPR-1B* (Q249R) for increase in ovulation rate in sheep.
- *TSHR* mutation and *SH3RF2* deletion growth and food intake in chicken.
- *LCORL* and *NR6A1* gene mutations affecting body length traits in pigs.

Some of these traits have monogenic basis while others have complex multifactorial basis. On the other hand, some disease-causing mutations are also unintentionally accumulated in domestic animals due to selective breeding of other traits. The white spotting locus in Boxer and Bull terrier and dorsal hair ridge in Rhodesian ridgebacks are the two first successful examples of monogenic traits mapped using GWAS in dogs (Karlsson *et al.*, 2007; Salmon Hillbertz *et al.*, 2007).

On the other hand complex traits where not only several genetic factors are responsible for the disease but also several environmental factors are involved are more difficult to map. GWAS is one of the most efficient gene-mapping strategies available to map complex traits. The results presented in this thesis (paper III) describe an example of complex trait mapping of demodicosis disease (which is an inflammatory skin disease caused by *D. canis* parasite) in the dog breed Staffordshire Bull Terrier. Other examples of polygenic complex diseases include canine systemic lupus erythematosus (SLE)-related disease complex which is caused by multiple risk loci identified using GWAS (Wilbe *et al.*, 2010) and atopic dermatitis in German shepherd dogs, which is caused partly by a regulatory mutation in *PKP-2* gene encoding Plakophilin-2, a protein critical for maintaining an efficient skin barrier (Tengvall *et al.*, 2013).

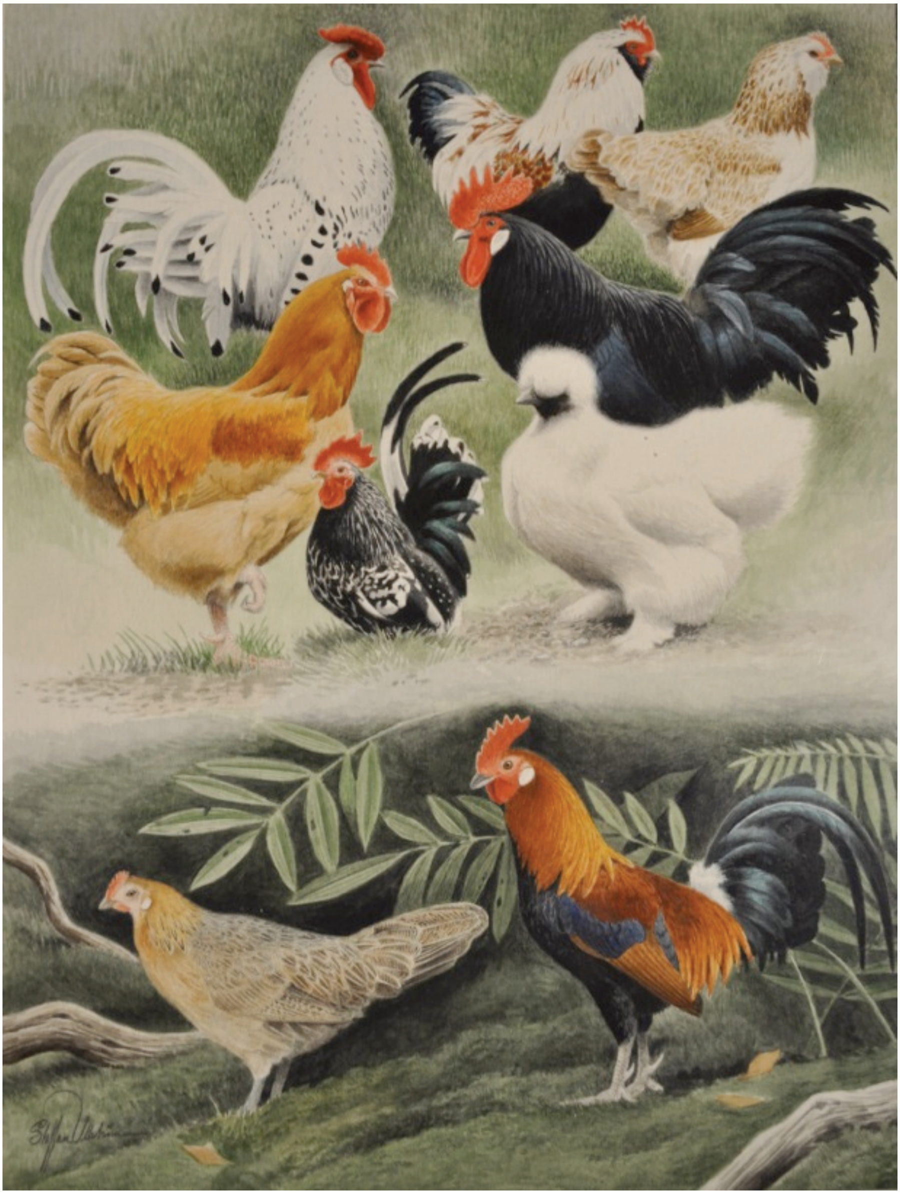

***Figure1.* The figure shows phenotypic diversity of different breeds of domestic chicken in comparison with wild ancestor Red Jungle Fowl** (Artist: Staffan Ullström)

1.4 Genome-wide association mapping

GWAS was used in paper III to perform an unbiased scan of the entire genome in healthy controls and cases affected by the demodicosis disease in Staffordshire Bull Terriers. The Canine Genome Project (Lindblad-Toh *et al.*, 2005) identified a large number of SNPs used to create canine SNP genotyping arrays that are used in GWAS (Karlsson *et al.*, 2007) and currently the high density 173K-genotyping array (Vaysse *et al.*, 2011) is also available, which was used in study III. In GWAS, the whole genome is scanned in all individuals from case and control population to identify the region of association. The associated haplotype contains regions where the healthy controls are genetically different from the affected individuals.

In GWAS in dogs, fewer markers (>15,000 SNPs) are needed as compared to humans (300,000 to 1,000,000 SNPs) (Gabriel *et al.*, 2002) due to the long haplotype blocks (Lindblad-Toh *et al.*, 2005).

1.5 Next Generation Sequencing methods

NGS also known as high-throughput sequencing, is a common term that describes several modern sequencing technologies. The most common techniques are Roche 454, Illumina (Solexa), Ion torrent, Ion Proton, PacBio and SOLiD. High-throughput sequencing technologies parallelize the sequencing process, producing thousands or millions of sequences at once. NGS technologies both in whole-genome sequencing (WGS) and whole-exome sequencing (WES) has not only lowered the cost of DNA sequencing as compared to standard Sanger sequencing methods but also provides unbiased approach for detecting large number of single nucleotide variants (SNV) which includes SNP and INDEL (insertions or deletions) and large structural variations (SV).

The most popular sequencing platform of choice has been the Illumina HiSeq Platform, which uses reversible terminator chemistry and optical modules to detect the fluorescent signal (Bentley *et al.*, 2008). However, there are other emerging technologies that provide an alternative choice for WGS and WES. One of those is the Ion Proton™ Platform (Merriman *et al.*, 2012) which uses semiconductor technology to generate sufficient amount of high-quality sequence data to cover large eukaryotic genomes in a relatively short time (paper IV).

1.6 Bioinformatic Methods

With the emergence of NGS technologies, that are producing huge amount of data, bioinformatic tools and statistical methods are needed to manage and analyze this data. Although several tools are already available for analysis of NGS data from very basic steps of raw data quality check and preprocessing (FASTQC), alignment to the reference genome (BWA, Mosaik) to the more complex downstream analysis of variants calling tools like SAMtools (Li *et al.*, 2009b) and GATK (for calling SNPs and INDELs) and for large structural variants (Magi *et al.*, 2010).

1.6.1 SNV functional analysis

Here I have summarized few methods and resources available to predict functional significance of SNPs and INDELs (paper I, II) detected using NGS methods. Figure 2 shows few standard tools that can be used to find functional significance of SNPs and INDELs in order to define causative mutations. The obtained SNVs from the GATK or SAMtools after quality filtering are used as an input to ANNOVAR software. ANNOVAR is a method that categorizes the SNVs into coding or non-coding variants using the gene models that the user provides (Figure 2). For examples in Paper I and Paper II, we used Ensembl gene models for annotation of the SNVs obtained from resequencing data of pigs and rabbits. The ANNOVAR software (Wang *et al.*, 2010) provides several other functionalities for finding if the SNV is overlapping some known transcription factor binding sites, histone marks or conserved sites.

PolyPhen-2 v2.2.2 (Polymorphism Phenotyping) is one of the tools that is widely used for evaluating missense SNPs. It predicts possible impact of an amino acid substitution on the structure and function of proteins. PolyPhen-2 predicts damaging or benign effects of non-synonymous variants based on eight sequence-based and three structure-based predictive features (Adzhubei *et al.*, 2010). An alternative tool to PolyPhen-2 is SIFT v4.0.5 (Sim *et al.*, 2012). SIFT predicts whether an amino acid substitution affects protein function and is based on sequence similarity and the physical properties of amino acids.

SIFT can be applied to naturally occurring non-synonymous polymorphisms and laboratory-induced missense mutations. The underlying SIFT algorithm is based on evolutionary conservation of the amino acids within protein families.

Although the PolyPhen-2 and SIFT prediction scores are positively correlated overall, they can be substantially different from each other, quantitatively as well as qualitatively. As a result, the SIFT-based and the PolyPhen-2-based results can also differ. These tools have recently added the functionality of predicting the effect of INDELs on protein function. SIFT and Polyphen-2 are the most widely used softwares due to their accuracy and sensitivity for prediction. Apart from SIFT and Polyphen-2 there are currently other programs also available that provide similar functionality *e.g.* HOPE, whereas PROVEAN and SIFT Indel are used for INDELs (Hu & Ng, 2013; Choi *et al.*, 2012).

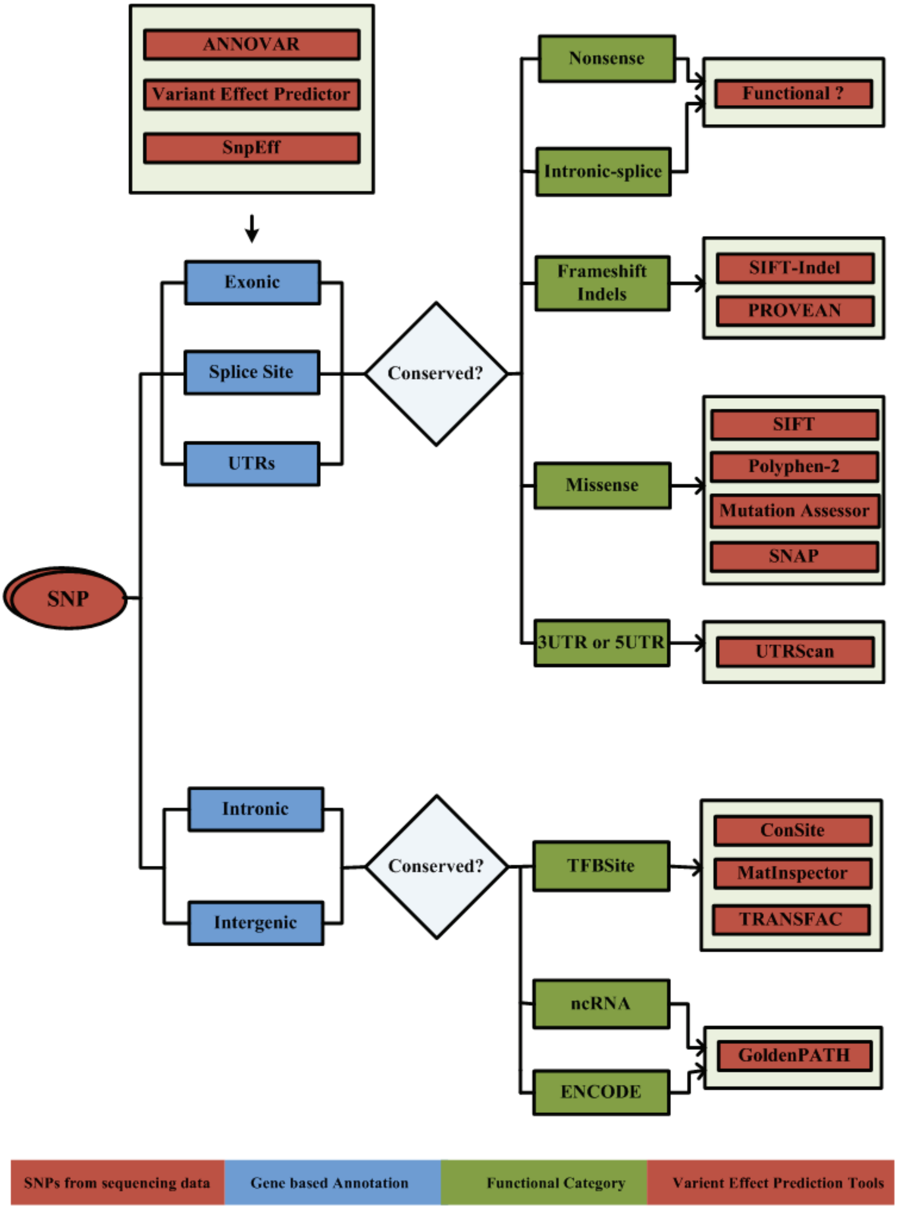

Figure2. Bioinformatics general pipeline for functional analysis of SNVs.

1.6.2 Structural Variants

Structural variants are the polymorphisms that change the structure of the genome either affecting the copy count of any genomic region, called copy number variants (CNVs), including insertions, deletions and duplications or variants that do not affect the copy count but alter the genomic region like inversions and translocations (paper V). The average size of structural variants (SV) is ~8 kb whereas the CNVs range from 50 bp to large structural events (Alkan *et al.*, 2011). Together with SNP, copy number changes become an important source of genetic variations either contributing to Mendelian traits or some of the missing heritability for complex traits (Salmon Hillbertz *et al.*, 2007; Giuffra *et al.*, 2002). In the past, several studies indicated that SVs have been associated with variety of human diseases (Yang *et al.*, 2013; McCarroll & Altshuler, 2007; Stranger *et al.*, 2007). Similar to in livestock genomes, research in genome-wide CNV identification of various domestic animals including dog (Axelsson *et al.*, 2013; Alvarez & Akey, 2012; Olsson *et al.*, 2011; Chen *et al.*, 2009), sheep (Fontanesi *et al.*, 2011), rabbit (Carneiro *et al.*, 2014; Fontanesi *et al.*, 2012), chicken (Jia *et al.*, 2013; Imsland *et al.*, 2012; Rubin *et al.*, 2010) and cattle (Bickhart *et al.*, 2012) showed their importance in genetic diversity and evolution (Molin *et al.*, 2014; Berglund *et al.*, 2012).

2 Aims of this Thesis

The overall aim of this thesis was to screen for candidate causative mutations underlying phenotypic changes in domestic animals using bioinformatics pipelines and methods. The datasets used for bioinformatics analysis in the current thesis was either generated by NGS techniques of whole genome resequencing or by genome-wide SNP arrays. The specific aims were as follows:

- Identify causative mutations for important phenotypic traits in domestic animals

- Sorting intolerant from tolerant mutations and loss of function mutations

- Building bioinformatics pipelines and using systems to manage and analyze NGS and GWAS data.

- Building methodology for finding putative structural variants.

3 Present Studies

3.1 Phenotypic changes during domestication (Papers I and II)

3.1.1 Background

Animal domestication not only results in significant changes in the morphology, reproduction, behavior and physiology of the animals but it also shapes their genome. Several of the main farm animals (cattle, sheep, goat and pig) were domesticated 9,000–11,000 years ago. Both European and Asian subspecies of wild boar have contributed to the European and Asian breeds of domestic pigs (Giuffra *et al.*, 2000). Due to the strong selective breeding for thousand of years, it has resulted in accumulation of new mutations with favored phenotypic effects. These mutations that modify the gene function or gene expression dominate as compared to the ones with pathological consequences that are eliminated by purifying selection.

For instance, at least for the last 200 years, breeders have selected for coat color diversity. It was shown in past that mutation in *MC1R* that encodes melanocortin receptor 1, controls the red and black pigment in many animal species including cattle, horse, pig and dog (Newton *et al.*, 2000; Kijas *et al.*, 1998; Marklund *et al.*, 1996; Klungland *et al.*, 1995). The dominant white phenotype in domestic pigs is caused by two mutations in the *KIT* gene encoding the mast/stem cell growth factor receptor (Marklund *et al.*, 1998). Similarly, strong selection for lean growth with high protein and low fat content was another trait that a breeder has selected for. In pigs and other species, a missense mutation (g.1843C>T) in Ryanodine receptor 1 (*RYR1*) is responsible for lean meat and muscularity (Fujii *et al.*, 1991). A regulatory mutation in intron 3 of insulin-like growth factor 2 (*IGF2*) gene was responsible for the increase in muscle growth and heart size and decrease in fat deposition in pigs (Van Laere *et al.*, 2003).

Rabbit domestication is a recent event which started ~1400 years ago in southern France. Iberian Peninsula and Southern France were populated with only two subtypes *Oryctolagus cuniculus cuniculus* and *O. c. algirus*. Interestingly, they are still populated with the same wild rabbit populations, making these the ancestors of domestic rabbits (Carneiro *et al.*, 2011). These facts offered a major advantage to carry out this study (Paper II) to infer the genetic basis of animal domestication as compared to many other domesticated species. Generation of a high-quality rabbit genome sequence made it possible to compare the wild and domesticated rabbit populations.

3.1.2 Results and discussion Paper I

In paper I, we performed whole genome resequencing to search for selective sweeps and genetic variants that showed marked allele frequency differences between domestic pig and wild boar populations. Here we sequenced eight different pools of pigs and wild boars at an average coverage of ~5x/pool. We did whole genome resequencing of pools of European domestic pigs and wild boars using SOLiD mate pair library (average insert of 1.3kb). The reads were mapped to the reference genome assembly Sscrofa10.2. After initial quality control and filtering, about 6.7 million SNPs were retained, that were used in the downstream analysis. These SNPs were further used for sweep analysis (beneficial genetic variants increase in frequency due to positive selection together with linked neutral sequence variants) in European domestic pigs by searching the regions with excess homozygosity, in which we scanned along the reference genome sequence in windows of size 150 kb. For each window, pooled Heterozygosity (Hp) and its Z-score (ZHp) was calculated (Rubin *et al.*, 2010). Windows with ZHp < -4 were retained as candidate sweep loci.

This approach revealed, 13 sweeps with ZHp<-5 and 64 loci with ZHp<-4. A major finding in this study was the striking correlation between putative sweep regions and well-established quantitative trait loci (QTL). Domestication leads to phenotypic changes (*e.g.* reproduction, physiology and morphology etc.), which was very well explained in this study. For example, a QTL for feed intake and growth lies in melanocortin 4 receptor (*MC4R*) gene locus (Kim *et al.*, 2000). One of the most striking finding in this study was the colocalization of three of the most convincing selective sweep candidates with major QTL that explains the elongation of the back and an increased number of vertebrae in domestic pigs (vertebrae 21-23) as compared to wild boar (vertebrae 19) (King & Roberts, 1960). Three of these sweeps were in genes: (***I***) Nuclear Receptor 6 A1 (*NR6A1*) that also contained a missense mutation (Pro192Leu) previously proposed to be the causative mutation (Mikawa *et al.*, 2007). (***II***) Pleomorphic adenoma gene 1 (*PLAG1*) that has been associated with variation in height in humans (Gudbjartsson *et al.*, 2008) as well as with a major QTL for height in cattle (Karim *et al.*, 2011). (***III***) Ligand-dependent nuclear receptor corepressor-like (*LCORL*) which has been associated with human stature and body size in dogs (Vaysse *et al.*, 2011) cattle (Pryce *et al.*, 2011) and horses (Signer-Hasler *et al.*, 2012). In order to verify some of these sweeps we overlapped them with extreme SNPs (allele frequency AF>0.9 and AF<0.1 in domestic pigs and wild boars, respectively) from individually sequenced (average coverage of 10x) domestic pigs (n=36) and wild boars (n=11).

We also searched for the coding SNPs that showed a marked allele frequency difference between European domestic pigs and wild boars (AF >80% in one group and AF < 20 % in the other group). These SNPs were further annotated using Ensembl gene models in ANNOVAR software (Wang *et al.*, 2010). Our results showed that gene inactivation did not play a prominent role during pig domestication, which was consistent with previous results in chickens (Rubin *et al.*, 2010). A complicating factor in the analysis of finding non-sense mutations that have become fixed or nearly fixed in domestic pigs was that the reference genome was obtained from a domestic pig, which means that if a nonsense mutation has become fixed in the domestic pig it is likely that the corresponding gene model may be wrong. In order to remove this bias we developed an alternative approach (Figure 3).

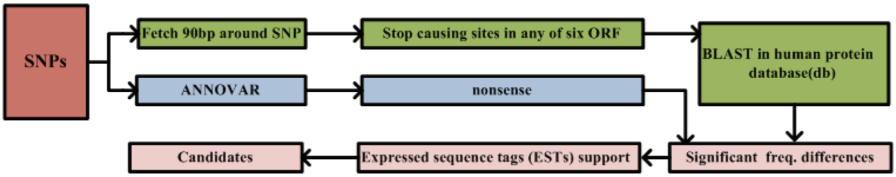
SNPs
Fetch 90bp around SNP
Stop causing sites in any of six ORF
BLAST in human protein database(db)
ANNOVAR
nonsense
Significant freq. differences
Expressed sequence tags (ESTs) support
Candidates

Figure 3. Pipeline for finding candidate nonsense SNPs. Blue and pink colors show general steps followed. Alternative approach (green and pink) for finding candidate nonsense mutations fixed in domestic pigs.

We found significant excess of derived non-synonymous substitutions (P = 0.00016) in domestic pigs. However, only 3 out of 72 missense SNPs overlapped sweeps showing that these SNPs have not increased in frequency due to recent sweeps. SIFT and Polyphen-2 (Sim *et al.*, 2012) analysis showed that 32 out of 72 were potentially damaging. Similarly, missense mutations in the genes corticosteroid-binding globulin (*CBG*) also known as *SERPINA6* affecting cortisol-binding capacity, hexokinase 2 (*HK2*) a key enzyme for glucose matabolism and Semaphorin 3D (*SEMA3D)* important in neuronal development are of particular interest.

We further identified four duplications (three of these within a large 450 kb duplication) at the *KIT* locus that were exclusively present in white or white-spotted pigs, carrying the *Dominant white*, *Patch*, or *Belt* alleles. We did not find a significant overlap of CNVs with our predicted sweeps showing that the majority of these behave as neutral markers. However, a duplication of 8 kb in the intron of Caspase 10 (*CASP10*) overlapped one of the major sweeps.

3.1.3 Results and discussion Paper II

The aim of this study was to understand how domestication leads to fixation of new mutations as compared to the selection of standing variations. Here we generated a female rabbit genome (OryCun2.0) assembly of size 2.66 Gb using Sanger sequencing. For annotation of the genome, we used Ensembl 73 gene models and RNA sequencing data for rabbit and human orthologs. To answering the domestication specific questions we performed whole genome resequencing of pooled samples (10x coverage) from domestic rabbits (6 breeds), wild rabbits (3 southern France and 11 Iberian Peninsula) and an outgroup snowshoe hare *(Lepus americanus)* and aligned these to the reference sequence.

Our results showed that rabbit is one of the most polymorphic mammals sequenced so far with approximately 50 million high quality SNPs and 5.6 million INDELs. The results from identity scores analysis confirms that domestic rabbits are more closely related to French wild rabbits and that there has been two bottlenecks, *i*) when rabbits from the Iberian Peninsula colonized southern France and *ii*) during domestication in Southern France. Using selective sweep analysis, we identified 78 significant sweeps. One of the sweeps was observed near glutamate receptor gene (*GRIK2*), which is highly expressed in brain and two sweeps near *SOX2* (*SRY-BOX2*) region, which encodes a transcription factor important for stem cell maintenance.

We also scanned for absolute allele frequency differences between domestic and wild rabbits (ΔAF= delta allele frequency), to find the extreme SNPs specific to each group "domestic or wild". Only 20 SNPs were completely fixed in domestic rabbits. Significant overrepresentation of high ΔAF SNPs in conserved non-coding elements was observed. An important finding was the nullification of an assumption that rapid evolutionary changes in rabbit domestication resulted due to gene inactivation. No nonsense or frame-shift mutation was observed in coding SNPs with high ΔAF that differ in wild and domesticated rabbits consistent with data from chicken (Rubin *et al*., 2010) and pigs as in paper I (Rubin *et al.*, 2012). We did not find any fixed missense mutations, but we identified 14 missense mutations with high ΔAF>0.90. We assume most of these missense mutations result from hitchhiking rather than being functional based on the observation of poor sequence conservation and similar chemical properties of substituted amino acids, its derived state of the domestic allele. However, two of the missense mutations stand out which may be candidate causative mutations. This includes tetratricopeptide repeat domain 21 B (*TTC21B* with a Q813R mutation) where all domestic rabbits shows amino acid Arginine while all wild rabbits were homozygous for the ancestral allele with amino acid Glutamine that was completely conserved in 55 mammals. The second example is the lysine-specific demethylase 6B (*KDM6B* with R1627W) where all wild rabbits and 44 mammals were completely conserved for Arginine.

Enrichment analysis showed that the most enriched biological process was "cell fate commitment" and statistical results supported brain and nervous system cell development, more than other categories suggesting brain and neuronal development have often been targeted during domestication. Extreme SNPs in non-coding regions overlapping conserved sites were associated with the following genes *BMP4*, *CTNNB1*, *EYA2*, *KLF4*, *PAX2*, *SIX2* and *SOX2*. Electrophoretic mobility shift assay (EMSA) with double-stranded oligonucleotide probes for transcription factors and nuclear extracts from mouse embryonic stem cell–derived neural stem cells showed, 7 out of 17 probes at *SOX2*, *KLF4* and *PAX2* with high ΔAF located at conserved sites showed differences in DNA-protein binding capacity between genotypes.

Our results suggest that genes affecting brain and neuronal development have often been targeted during domestication. We observed shifts in allele frequencies rather than complete fixation of causative mutations. We also observed that changes in non-coding sequences are numerously much more important than changes in coding sequences during rabbit domestication. We also propose that a single specific "domestication gene" may not exist, because tameness has a highly polygenic background and evolved by shifts in allele frequencies at many loci, rather than by critical changes at only a few 'domestication loci'.

3.2 GWAS identifies candidate risk loci for demodicosis in Staffordshire Bull terrier (Paper III)

3.2.1 Background

Canine demodicosis is an inflammatory parasitic skin disease caused by acarine mites of *Demodex canis* that are present and proliferate in canine hair follicles causing cutaneous lesions (Chesney, 1999). The disease is prevelant in several breeds including American Staffordshire bull terrier, Staffordshire Bull Terrier and Chinese shar-pei (Plant *et al.*, 2011). The Staffordshire Bull Terrier is a high-risk breed for demodicosis with an odds ratio of 17.1 (Plant *et al.*, 2011). Previous studies have shown association of demodicosis to the canine MHC class II region (Dog Leukocyte Antigen, DLA) (It *et al.*, 2010). The disease is sub-divided into juvenile and adult demodicosis based on age of onset. The juvenile form (onset prior to 18 months of age) is associated with a strong breed and family predilection and has been assumed to have a genetic background (Miller *et al.*, 1992) whereas the adult form appears in old dogs (>18 months of age) often in association to an immune compromising disease or immune-suppressive treatment (Duclos *et al.*, 1994; Miller *et al.*, 1993)

The juvenile-onset demodicosis can further be sub-divided based on severity of the disease into localized (with less than 3-5 areas of body affected) and generalized (greater than 5 areas or whole body is affected).

3.2.2 Results and discussion Paper III

In this study we used 262 individuals (198 Swedish dogs and 64 Norwegian dogs) of which 113 were cases (affected by juvenile-onset demodicosis with either generalized type (n=61) or localized type (n=52)) and 149 were controls (unaffected). All individuals were genotyped using 170K Illumina HD canine SNP array. For each individual dog, the phenotype (case and control) was carefully characterized. Dogs with clinical signs compatible with demodicosis with an onset prior to 18 months age, and where the diagnosis was confirmed with direct microscopy were classified as cases.

The cases were further sub-divided into localized (<3-5 small areas affected) versus generalized (>5 areas affected or widespread disease) in accordance to standard classification. Healthy controls were >3 years old, never having had any evidence of alopecia. The Swedish dogs included as healthy controls, had skin scrapings and trichogram taken from three randomly chosen areas, revealing no Demodex mites at direct microscopy. For each dog, gender, the age of onset, and classification (case localized, case generalized or control) was recorded. The relationship between the individuals was characterized and individuals with relatedness closer than at grandparental level were removed. However, for Norwegian dogs, we had either cases with whole body affected (n=8) or more than 6 spots affected (n=9). In the downstream analysis we merged whole body affected Norwegian dogs with the Swedish generalized dogs to prepare a dataset mainly having generalized cases. Similarly other datasets were also prepared by using the disease severity criteria. Each dataset was analyzed separately.

After removal of outliers and quality control filtering we retained about 99,232 SNPs (~58% of markers) for association analysis. Demodicosis-gender relationship was found non significant in each dataset using Fisher's exact test.

Our initial association analysis showed population stratification with λ (genomic inflation factor) greater than 1.4 for each dataset. The demodicosis population showed two clusters when we performed the Kmean clustering. The stratification was due to the uneven distribution of cases and controls in two clusters. Population stratification and cryptic relatedness between the individuals (which is common in dog breeds) was successfully corrected for by using mixed model (with genomic kinship and binomial trait information) for each dataset with λ=0.99.

We identified four preliminary disease associated loci on CFA29, CFA28, CFA18 and CFA17. All four loci were candidates and none passed Bonferroni correction. For the dataset of generalized demodicosis only, we observed association on CFA29 with raw pvalue of 3.5×10^{-5}. The total associated region on chromosome 29 with the top SNP at 17,150,595 bp was defined as approximately 2 Mb long containing the genes (*MYBL1, VCPIP1, SGK3, MCMDC2, TCF24, PPP1R42, COPS5, CSPP1, ARFGEF1, CPA6,* and prostaglandin reductase 1 pseudogene*, PREX2*). We defined the associated haplotype using LD clumping using r2 =0.8 (Purcell *et al.*, 2007). The associated haplotype for CFA17 spanning 2 Mb contained the genes (*LRRTM1, CTNNA2* and *REG3A*). Whereas the haplotypes identified on CFA28 and CFA18 contain the genes (*GOT1, NKX2, SLC25A28, ENTPD7, COX15,* and *FGFR2*-like) and genes (*CCDC73, EIF3M, WT1, RCN1, ELP4, IMMP1L, DNAJC24,* and *DCDC1*), respectively. The associated regions contain several interesting candidate genes that will be further investigated for their function. For example, on CFA17, the gene *REG3A* (regenerating islet-derived 3 alpha), which is bactericidal C-type lectin that acts against gram-positive bacteria and mediate their killing, it also regulates keratinocyte proliferation and differentiation after skin injury via activation of the EXTL3-AKT signaling pathway. Another candidate gene was Fibroblast growth factor receptor *FGFR2*-like locus on CFA28 which was about 100 kb downstream to top candidate SNP. The gene *FGFR2* has been involved in controlling the epidermal barrier and cutaneous homeostasis in keratinocytes (Yang *et al.*, 2010).

This study identified some preliminary candidate associated regions (where none of the loci passed through Bonferroni correction). Further evaluation of some of the associated regions containing interesting candidate genes will be of particular interest either by doing permutation testing using random sampling of phenotypes or by adding more samples to determine if that gains additional power to detect association and ultimately identify the candidate causative mutation and other genetic risk factors.

3.3 Methodologies for detection of SNV and large structural variants using whole genome resequencing (Papers IV and V)

3.3.1 Background

Whole genome sequencing (WGS) of individual genomes with NGS technologies has triggered numerous groundbreaking discoveries and ignited a revolution in genomic science. It has opened a new avenue for personalized healthcare and medicine based on the detection of genetic variations related to disease. These have not only reduced cost of sequencing individual genomes, but also provides powerful and unbiased (Boycott *et al.*, 2013) approach for detecting larger proportion of genetic variation, from single base pair changes, INDELs, structural variants, chimeric transcripts and gene rearrangements affecting phenotype.

Illumina HiSeq sequencing platform dominates the sequencing market today, which uses reversible terminator chemistry and optical modules to detect the fluorescent signal (Bentley *et al.*, 2008). However, there are other emerging technologies that provide an alternative choice for WGS and WES. One of those is Ion Proton™ Platform (Merriman *et al.*, 2012) which uses semiconductor technology to generate sufficient amount of data to cover large eukaryotic genomes in a relatively short time.

Previously, the structural variants were discovered by either whole genome array comparative genome hybridization (aCGH) in which the relative frequencies of probe DNA segments between two genomes was compared (Pinkel *et al*., 1998) or using Hapmap available data and SNP arrays measuring the intensities of probe signals at known SNP loci (International HapMap *et al.*, 2010). Multiple Ligation dependent Genome Amplification was also used for their detection (Salmon Hillbertz *et al.*, 2007). Sanger sequencing of paired reads was used as an alternative to the above-mentioned methods to detect CNVs, inversions and translocations with high accuracy and resolution at the expense of time and cost. Today several methods have been developed using the NGS data to detect the SVs with each offering some limitations.

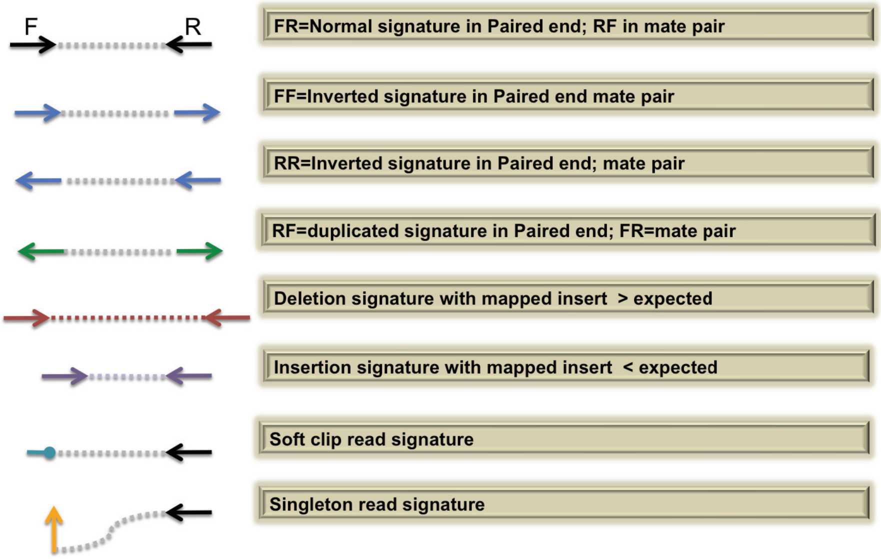
F
R
FR=Normal signature in Paired end; RF in mate pair
FF=Inverted signature in Paired end mate pair
RR=Inverted signature in Paired end; mate pair
RF=duplicated signature in Paired end; FR=mate pair
Deletion signature with mapped insert > expected
Insertion signature with mapped insert < expected
Soft clip read signature
Singleton read signature

Figure 4. **Description of read signatures in paired end mapping (PEM) and mate pair (MP). Forward read market as F and reverse read marked as R**.

In general, there are four categories of methods to detect SVs using NGS data (Alkan *et al.*, 2011);

1. Depth of Coverage (DoC) methods
2. Paired end mapping (PEM) methods
3. Split read methods (SR) and
4. Assembly (AS) based methods

The assumption of DoC-based methods (*e.g.* CNVseq and CNVnator) is that the coverage is uniform *i.e.* the number of reads mapped to a region are assumed to follow the Poisson distribution, however there is bias due to GC-content and mapability and these methods are unable to detect inversions and translocations (Xie *et al.*, 2009; Abyzov *et al.*, 2011). PEM methods (*e.g.* Breakdancer) use the information of paired reads and their orientation (Figure 4, Paper V). The insert size is used to detect the insertions and deletions, although the size of CNVs detected is limited by the size of insert in this method (Xi *et al.*, 2011; Chen *et al.*, 2009; (Rausch *et al.*, 2012). SR methods (*e.g.* Pindel) use the information of anchored reads to identify the breakpoint locations while AS methods (*e.g.* SOAPdenovo) are based on the *de novo* assembly (Ye *et al.*, 2009; Li *et al.*, 2010). Today, several tools have been developed that use combination of methods like PEM and DoC information (GenomeSTRiP, SVDetect) to detect SVs (Handsaker *et al*., 2011; Zeitouni *et al*., 2010).

3.3.2 Results and discussion Paper IV

In paper IV we describe results from resequencing of four Chinese Crested dogs using the Ion Proton platform and evaluated the platform for whole genome resequencing of the canine genome for coverage (genome and exome), genotype concordance with SNP array and number of variants detected. For each dog, we constructed a 200 bp fragment library that was sequenced on two Ion PI^{TM} Chips with 500 run flows. Each chip generated about 9.5 Gb sequence data having approximately 73.3 million single reads with mean read length of 130 bp. The reads were aligned to the canine reference genome sequence (CanFam3.1). On average 98% of the reads could be aligned to the reference genome with 8x coverage. However, after removal of duplicates (21%), we obtained on average 6x coverage for the autosomal genome and 236x coverage for mitochondrial genome. Our results revealed that on average 80% of whole genome and 77% of exome (using Ensembl transcripts) were covered with at least 4 reads per base. The read coverage over the canine genome with respect to GC content showed normal coverage at approximately 40% GC content. However, we detected a gradual drop in coverage if the GC content was less than 20% or greater than 60%. We also observed fluctuation of the mean base quality at GC content greater than 80%.

Variants (SNPs and INDELs) were called on all four dogs together (combined analysis) and each dog separately (individual dog) using both SAMtools (Li *et al.*, 2009a) and GATK (Van der Auwera *et al.*, 2002). After hard filtering, the combined analysis resulted in ~10 million filtered variants using SAMtools and ~7 million filtered variants using GATK UnifiedfiedGenotyper. However, the individual analysis generated about 3 million variants per individual dog. Previous re-sequencing studies report 6.1 to 7.4 million variants detected per dog (Drogemuller *et al.*, 2014; Gilliam *et al.*, 2014; Guo *et al.*) with depth of coverage from 20X to 34X. This indicates that there is a trade-off among number of individuals sequenced and sequencing depth in terms of detected variation. In terms of tool differences, number of detected INDELs using SAMtools almost two fold the number of UnifiedGenotyper detected INDELs, whereas the number did not increase when we did the overlap with known variants. This might be an indication of high false positive rate of SAMtools even after variant filtration.

In order to quality control the genotype calls, we genotyped two of the sequenced individuals with Illumina HD Canine SNP array comprising 174

037 markers. Our results showed that with the existing coverage we could achieve more than 90% genotype concordance for both the individuals. However, the majority of the mismatches observed in discordant genotypes (7%) were due to SNP positions where the individual had been called as homozygous for the reference type allele by UnifiedGenotyper, but heterozygous by the SNP array.

We further evaluated how increase in coverage and number of prepared libraries from the same sample could potentially increase sufficient coverage of regions. We did a library merging simulation, and our results showed that when we combined two libraries, the proportion of coverage increased up to 94.6% genome-wide and up to 90.8 % exome-wide in comparison to coverage obtained using a single library. The results showed that decrease in uncovered area and low covered area of genome would be the highest when sequencing two libraries per individual each on two chips, However, even after merging all four available libraries, there was still around 1% of genome and almost 3% of exome that remained uncovered. We found that most of the uncovered regions were overlapping repeat regions with high GC content which is one of the common problem in PCR-based sequencing platforms (Aird *et al.*, 2011). After eliminating these most common lack of coverage issues there was still a small fraction left that has no obvious explanation and might be interesting from a biological point of view.

3.3.3 Results and discussion Paper V

Here we have developed a tool based on deviant read and read orientation named as "DevRO" to identify candidate structural variants putatively in multiple populations. We have used paired-end mapping (PEM) method for identification of inversions and both paired-end mapping method followed by depth of coverage (DoC) approach to screen for candidate deletions and duplications. DevRO can detect SVs in multiple populations without doing pairwise comparisons and using combined approach (PEM and DoC) that gives power to the study as compared to traditional methods that are based either on PEM or DoC. DevRO is also able to detect deletions in the reference assembly, which is an added functionality as compared to the available methods.

DevRO was implemented in perl, v5.10.0 and needs as an input raw binary alignment mapping files (BAM) from NGS technologies (paired end or mate pair data). There are three main modules of DevRO; 1) VariantCaller 2) VariantParser and 3) VariantAnnotate.

DevRO Variantcaller was used to call the raw variants on multiple populations. Here we scanned whole genome for PEM signatures in windows of 1 kb. For each locus we stored the information of deviant reads in each population. This information was further processed at the VariantParser to calculate the fraction of deviant reads in two groups in order to find SVs with significant frequency differences between the two groups (as test case we used data from domestic and wild Rabbits). We further used VariantAnnotate to score the variants using DoC for deletion and duplications, gene annotations.

As a test case we have used Rabbit mate pair (MP) data (3x coverage, and 4.5 kb insert) generated from two wild and two domestic rabbits for PEM analysis and Rabbit paired-end sequencing data from pooled samples of wild and domestic rabbit populations with 10x coverage/population (Carneiro *et al.*, 2014; Paper II). The reads were preprocessed and aligned to the rabbit reference genome sequence (OryCun2.0) using BWA.

To check for concordance we used SVDetect (Zeitouni *et al.*, 2010) which is an already available method using both PEM and DoC information for detecting the SVs in multiple populations, Our results showed that number of inversions detected was 90 and 411 for SVDetect and DevRO, respectively, with 80 overlapping inversions using reciprocal fraction overlap of 0.7. One possible reason for detection of less number of inversions using SVDetect was the pairwise comparisons that are done between domestic and wild populations. The CNVs overlap between SVDetect and DevRO was 500 and 391 for deletions and duplications, respectively.

4 General Discussion and Future Prospects

Paper I and II have described some of the phenotypic changes that have occurred during the domestication process. Paper I highlights the changes in the body length and number of vertebrae in domestic pigs as compared to wild ancestors. This is very logical due to the strong selection in domestic pigs for meat production. Paper II sheds light on many genes affecting brain and neuronal development in domestic rabbits that have been under strong selection clearly showing that behavioral changes during the initial steps in animal domestication have allowed the domestic animals to live in the human environment. There are some main differences between the two studies that were of advantage. In rabbits the fairly recent history (~1400 years) and availability and well-defined origin of wild ancestors provides unique opportunity to study the domestication processes that allowed the animals to tolerate the human pressures as compared to pigs (10,000 years).

One of the complexities that arise in domestic animals, when we have a reference assembly from the domestic individual, and which makes it difficult to identify events like fixed deletions in domestic animals, is that we do not have the complete sequence for wild individuals. This could possibly be improved by making a second genome assembly available for a wild ancestor of each domestic animal. Better gene models are very important in functional annotation of the variants, as the Broad annotations are now being available for most of the species, with support from RNAseq data, it will be useful to adopt available softwares for these annotations which will enable better predictions of functions as compared to Ensembl annotations with many erroneous gene models.

In paper I and II we identified few candidate causative mutations using SIFT and Polyphen-2 softwares. These software predicts, if the amino acid substitution affect the protein function, by sorting the candidates into benign or possibly damaging or probably damaging categories. One of the candidate identified was *KDM6B* (lysine-specific demethylase 6B) in rabbits and hexokinase 2 (*HK2*) in pigs that showed marked allele frequency difference between domestic and wild rabbits and pigs, respectively. In future, the functional effect of missense SNPs specially those overlapping the selective sweep regions could be studied using cell culture methods by overexpressing the genes with the mutation and study the functional consequences.

Paper III shows the complexity of disease when dealing with a common disease where both environmental factors and several risk loci could potentially be involved in disease association and development. Our preliminary results show lack of significant association at the candidate loci in all of our analysis for the demodicosis disease. We believe, this could be due to several factors that could influence our results. Here are some of the possible reasons; (*i*) The lack of significant association in many GWAS of complex traits may be due to the Fisher's infinitesimal model that assumes a very large number of loci with each of them having small effects. (*ii*) We believe that the Bonferroni threshold using the number of SNPs, is a too conservative approach since many of the SNPs are linked to at least one neighbouring SNP and are not independent. An alternative to this approach in future for the current study would be to use permutation test using random sampling of phenotypes (Karlsson *et al.*, 2013). (*iii*) Population stratification is another factor that might contribute to the identification of false positives. We observed inflation factor $\lambda >=1.4$, this was explained by the non-homogeneous distribution of cases and controls in subpopulations identified using Kmeans clustering. This might have contributed to the population substructures. The association analysis for each of the subpopulations would be of particular interest to see if there is significant difference between the subpopulations for the prevalence of disease. (*iv*) Small mutation effect size could possibly be improved by increasing the number of subjects. (*v*) Density of informative markers, the degree of LD and the number of genotyped markers are some of the factors that might affect the results from GWAS. However, increasing the number of subjects if the mutation effect size is small could control several of these factors.

In the future, another interesting scenario would be to perform whole genome resequencing of few generalized cases and few controls at a higher depth. In order to focus on the associated regions obtained from GWAS results.

Currently, in this study we have used Staffordshire Bull Terriers (odds ratio for the demodicosis is 17.1). It would be interesting to add more samples from other high-risk breeds like American Staffordshire bull terrier (odds ratio for disease is 35.6) and Chinese shar-pei (7.2) (Plant *et al.*, 2011).

As the NGS technologies are becoming more and more common with low costs and less time for sequencing, in the future one could perform whole genome resequencing of a few cases and controls if only a few cases are available if one can assume a monogenic origin of the disease.

For detection of SNV and SV there is already a large number of online and standalone bioinformatics tools available (Paper IV and V). In the future, these tools could be further improved to make them species and platform independent. In Paper V we developed a tool DevRO for detection of structural variations using mate pair and paired-end sequencing data in multiple populations. DevRO uses combination of available methods (PEM and DoC) and predicts the SVs with significant frequency differences between the two groups. It also provides a unique feature for detecting deletions present in the reference assembly. Such deletions may be due to assembly errors or because the individual used for the assembly carry one or more deletions. In future, the precise detection of breakpoints using base quality of deviant reads and supplement to some routines with the *de novo* assembly might improve its prediction. A better visualization tool (user interface) for SV could be one option to improve DevRO and many other available tools.

5 Conclusion

The results of this thesis highlight some of the candidate causative mutations using bioinformatics pipelines, NGS and GWAS data. The main conclusions were as follows:

- Highlights phenotypic changes during domestication, by revealing sweeps in *NR6A1*, *LCORL* and *PLAG1* genes that underlie a major QTL for body length and increased number of vertebrae in domestic pigs. It also sheds light on the understanding of how alleles in domestic animals evolve by accumulation of multiple causative mutations.

- Gene inactivation has not contributed to the rapid evolution of domestic pigs and rabbits. Most of the derived mutations are regulatory influencing gene expression.

- Genes affecting brain and neuronal development (like *SOX2* and *GRIK2*) have often been targeted during domestication. We proposed that "single domestication genes" may not exist, and that tameness evolved by shifts in allele frequencies at many loci, rather than by critical changes at only a few 'domestication loci'.

- Genome-wide association mapping identified preliminary associated regions on CFA17, 18, 28 and 29 containing several interesting candidate genes that may increase knowledge about risk factors contributing to juvenile-onset demodicosis disease in Staffordshire Bull Terriers. As an example, the *REG3A* and *FGFR2l* genes are of particular interest.

- The thesis also highlights some opportunities and pitfalls of Ion Proton™ Platform in whole genome re-sequencing of the dog genome. An alternative approach for genome-wide association studies when monogenic diseases with autosomal recessive inheritance are investigated and small sample sizes are available.

- The bioinformatics pipeline “DevRO” for detection of structural variants (deletions, duplications, inversions and deletions in the reference genome) will be useful when large numbers of populations are re-sequenced as compared to traditional methods for detection of structural variants in a pairwise manner.

References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. & Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods,* 7(4), pp. 248-9.

Aird, D., Ross, M.G., Chen, W.S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C. & Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol,* 12(2), p. R18.

Alkan, C., Coe, B.P. & Eichler, E.E. (2011). Genome structural variation discovery and genotyping. *Nat Rev Genet,* 12(5), pp. 363-76.

Alvarez, C.E. & Akey, J.M. (2012). Copy number variation in the domestic dog. *Mamm Genome,* 23(1-2), pp. 144-63.

Andersson, L. (2001). Genetic dissection of phenotypic diversity in farm animals. *Nat Rev Genet,* 2(2), pp. 130-8.

Andersson, L. (2009). Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. *Genetica,* 136(2), pp. 341-9.

Axelsson, E., Ratnakumar, A., Arendt, M.L., Maqbool, K., Webster, M.T., Perloski, M., Liberg, O., Arnemo, J.M., Hedhammar, A. & Lindblad-Toh, K. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature,* 495(7441), pp. 360-4.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., Carter, R.J., Keira Cheetham, R., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., Irving, L.J., Karbelashvili, M.S., Kirk, S.M., Li, H., Liu, X., Maisinger, K.S., Murray, L.J., Obradovic, B., Ost, T., Parkinson, M.L., Pratt, M.R., Rasolonjatovo, I.M., Reed, M.T., Rigatti, R., Rodighiero, C., Ross, M.T., Sabot, A., Sankar, S.V., Scally, A., Schroth, G.P., Smith, M.E., Smith, V.P., Spiridou, A., Torrance, P.E., Tzonev, S.S., Vermaas, E.H., Walter, K., Wu, X., Zhang, L., Alam, M.D., Anastasi, C., Aniebo, I.C., Bailey, D.M., Bancarz, I.R., Banerjee, S., Barbour, S.G., Baybayan, P.A., Benoit, V.A., Benson, K.F., Bevis, C., Black, P.J., Boodhun, A., Brennan, J.S., Bridgham, J.A., Brown, R.C., Brown, A.A., Buermann, D.H., Bundu, A.A., Burrows, J.C., Carter, N.P., Castillo, N., Chiara, E.C.M., Chang, S., Neil Cooley, R., Crake, N.R., Dada, O.O., Diakoumakos, K.D., Dominguez-Fernandez, B., Earnshaw, D.J., Egbujor, U.C., Elmore, D.W., Etchin, S.S., Ewan, M.R., Fedurco, M., Fraser, L.J., Fuentes Fajardo, K.V., Scott Furey, W., George, D., Gietzen, K.J., Goddard, C.P., Golda, G.S., Granieri, P.A., Green, D.E., Gustafson, D.L., Hansen, N.F., Harnish, K., Haudenschild, C.D., Heyer, N.I., Hims, M.M., Ho, J.T., Horgan, A.M., Hoschler, K., Hurwitz, S., Ivanov, D.V., Johnson, M.Q., James, T., Huw Jones, T.A., Kang, G.D., Kerelska, T.H., Kersey, A.D., Khrebtukova, I., Kindwall, A.P., Kingsbury, Z., Kokko-Gonzales, P.I., Kumar, A., Laurent, M.A., Lawley, C.T., Lee, S.E., Lee, X., Liao, A.K., Loch, J.A., Lok, M., Luo, S., Mammen, R.M., Martin, J.W., McCauley, P.G., McNitt, P., Mehta, P., Moon, K.W., Mullens, J.W., Newington, T., Ning, Z., Ling Ng, B., Novo, S.M., O'Neill, M.J., Osborne, M.A., Osnowski, A., Ostadan, O., Paraschos, L.L., Pickering, L., Pike, A.C., Pike, A.C., Chris Pinkard, D., Pliskin, D.P., Podhasky, J., Quijano, V.J., Raczy, C., Rae, V.H., Rawlings, S.R., Chiva Rodriguez, A., Roe, P.M., Rogers, J., Rogert Bacigalupo, M.C., Romanov, N., Romieu, A., Roth, R.K., Rourke, N.J., Ruediger, S.T., Rusman, E., Sanches-Kuiper, R.M., Schenker, M.R., Seoane, J.M., Shaw, R.J., Shiver, M.K., Short, S.W., Sizto, N.L., Sluis, J.P., Smith, M.A., Ernest Sohna Sohna, J., Spence, E.J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C.L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S.M., Wakelin, S., Walcott, G.C., Wang, J., Worsley, G.J., Yan, J., Yau, L., Zuerlein, M., Rogers, J., Mullikin, J.C., Hurles, M.E., McCooke, N.J., West, J.S., Oaks, F.L., Lundberg, P.L., Klenerman, D., Durbin, R. & Smith, A.J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature,* 456(7218), pp. 53-9.

Berglund, J., Nevalainen, E.M., Molin, A.M., Perloski, M., Consortium, L., Andre, C., Zody, M.C., Sharpe, T., Hitte, C., Lindblad-Toh, K., Lohi, H. & Webster, M.T. (2012). Novel origins of copy number variation in the dog genome. *Genome Biol,* 13(8), p. R73.

Bickhart, D.M., Hou, Y., Schroeder, S.G., Alkan, C., Cardone, M.F., Matukumalli, L.K., Song, J., Schnabel, R.D., Ventura, M., Taylor, J.F., Garcia, J.F., Van Tassell, C.P., Sonstegard, T.S., Eichler, E.E. & Liu, G.E. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res,* 22(4), pp. 778-90.

Boycott, K.M., Vanstone, M.R., Bulman, D.E. & MacKenzie, A.E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet,* 14(10), pp. 681-691.

Carneiro, M., Afonso, S., Geraldes, A., Garreau, H., Bolet, G., Boucher, S., Tircazes, A., Queney, G., Nachman, M.W. & Ferrand, N. (2011). The genetic structure of domestic rabbits. *Mol Biol Evol,* 28(6), pp. 1801-16.

Carneiro, M., Rubin, C.J., Di Palma, F., Albert, F.W., Alfoldi, J., Barrio, A.M., Pielberg, G., Rafati, N., Sayyab, S., Turner-Maier, J., Younis, S., Afonso, S., Aken, B., Alves, J.M., Barrell, D., Bolet, G., Boucher, S., Burbano, H.A., Campos, R., Chang, J.L., Duranthon, V., Fontanesi, L., Garreau, H., Heiman, D., Johnson, J., Mage, R.G., Peng, Z., Queney, G., Rogel-Gaillard, C., Ruffier, M., Searle, S., Villafuerte, R., Xiong, A., Young, S., Forsberg-Nilsson, K., Good, J.M., Lander, E.S., Ferrand, N., Lindblad-Toh, K. & Andersson, L. (2014). Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science,* 345(6200), pp. 1074-9.

Chen, W.K., Swartz, J.D., Rush, L.J. & Alvarez, C.E. (2009). Mapping DNA structural variation in dogs. *Genome Res,* 19(3), pp. 500-9.

Chesney, C.J. (1999). Short form of Demodex species mite in the dog: occurrence and measurements. *J Small Anim Pract,* 40(2), pp. 58-61.

Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. & Chan, A.P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One,* 7(10), p. e46688.

Drogemuller, M., Jagannathan, V., Becker, D., Drogemuller, C., Schelling, C., Plassais, J., Kaerle, C., Dufaure de Citres, C., Thomas, A., Muller, E.J., Welle, M.M., Roosje, P. & Leeb, T. (2014). A mutation in the FAM83G gene in dogs with hereditary footpad hyperkeratosis (HFH). *PLoS Genet,* 10(5), p. e1004370.

Duclos, D.D., Jeffers, J.G. & Shanley, K.J. (1994). Prognosis for treatment of adult-onset demodicosis in dogs: 34 cases (1979-1990). *J Am Vet Med Assoc,* 204(4), pp. 616-9.

Fontanesi, L., Beretti, F., Martelli, P.L., Colombo, M., Dall'olio, S., Occidente, M., Portolano, B., Casadio, R., Matassino, D. & Russo, V. (2011). A first comparative map of copy number variations in the sheep genome. *Genomics,* 97(3), pp. 158-65.

Fontanesi, L., Martelli, P.L., Scotti, E., Russo, V., Rogel-Gaillard, C., Casadio, R. & Vernesi, C. (2012). Exploring copy number variation in the rabbit (Oryctolagus cuniculus) genome by array comparative genome hybridization. *Genomics,* 100(4), pp. 245-51.

Fujii, J., Otsu, K., Zorzato, F., de Leon, S., Khanna, V.K., Weiler, J.E., O'Brien, P.J. & MacLennan, D.H. (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science,* 253(5018), pp. 448-51.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J. & Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science,* 296(5576), pp. 2225-9.

Georges, M. (2007). Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. *Annu Rev Genomics Hum Genet,* 8, pp. 131-62.

Gilliam, D., O'Brien, D.P., Coates, J.R., Johnson, G.S., Johnson, G.C., Mhlanga-Mutangadura, T., Hansen, L., Taylor, J.F. & Schnabel, R.D. (2014). A Homozygous KCNJ10 Mutation in Jack Russell Terriers and Related Breeds with Spinocerebellar Ataxia with Myokymia, Seizures, or Both. *Journal of Veterinary Internal Medicine,* 28(3), pp. 871-877.

Giuffra, E., Kijas, J.M., Amarger, V., Carlborg, O., Jeon, J.T. & Andersson, L. (2000). The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics,* 154(4), pp. 1785-91.

Giuffra, E., Tornsten, A., Marklund, S., Bongcam-Rudloff, E., Chardon, P., Kijas, J.M., Anderson, S.I., Archibald, A.L. & Andersson, L. (2002). A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking KIT. *Mamm Genome,* 13(10), pp. 569-77.

Gudbjartsson, D.F., Walters, G.B., Thorleifsson, G., Stefansson, H., Halldorsson, B.V., Zusmanovich, P., Sulem, P., Thorlacius, S., Gylfason, A., Steinberg, S., Helgadottir, A., Ingason, A., Steinthorsdottir, V., Olafsdottir, E.J., Olafsdottir, G.H., Jonsson, T., Borch-Johnsen, K., Hansen, T., Andersen, G., Jorgensen, T., Pedersen, O., Aben, K.K., Witjes, J.A., Swinkels, D.W., Heijer, M.d., Franke, B., Verbeek, A.L.M., Becker, D.M., Yanek, L.R., Becker, L.C., Tryggvadottir, L., Rafnar, T., Gulcher, J., Kiemeney, L.A., Kong, A., Thorsteinsdottir, U. & Stefansson, K. (2008). Many sequence variants affecting diversity of adult human height. *Nat Genet,* 40(5), pp. 609-615.

Guo, J., Johnson, G.S., Brown, H.A., Provencher, M.L., da Costa, R.C., Mhlanga-Mutangadura, T., Taylor, J.F., Schnabel, R.D., O'Brien, D.P. & Katz, M.L. A CLN8 nonsense mutation in the whole genome sequence of a mixed breed dog with neuronal ceroid lipofuscinosis and Australian Shepherd ancestry. *Molecular Genetics and Metabolism*(0).

Hu, J. & Ng, P.C. (2013). SIFT Indel: predictions for the functional effects of amino acid insertions/deletions in proteins. *PLoS One,* 8(10), p. e77940.

Hunter, D.J. (2005). Gene-environment interactions in human diseases. *Nat Rev Genet,* 6(4), pp. 287-98.

Imsland, F., Feng, C., Boije, H., Bed'hom, B., Fillon, V., Dorshorst, B., Rubin, C.J., Liu, R., Gao, Y., Gu, X., Wang, Y., Gourichon, D., Zody, M.C., Zecchin, W., Vieaud, A., Tixier-Boichard, M., Hu, X., Hallbook, F., Li, N. & Andersson, L. (2012). The Rose-comb mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. *PLoS Genet,* 8(6), p. e1002775.

International HapMap, C., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Peltonen, L., Dermitzakis, E., Bonnen, P.E., Altshuler, D.M., Gibbs, R.A., de Bakker, P.I., Deloukas, P., Gabriel, S.B., Gwilliam, R., Hunt, S., Inouye, M., Jia, X., Palotie, A., Parkin, M., Whittaker, P., Yu, F., Chang, K., Hawes, A., Lewis, L.R., Ren, Y., Wheeler, D., Gibbs, R.A., Muzny, D.M., Barnes, C., Darvishi, K., Hurles, M., Korn, J.M., Kristiansson, K., Lee, C., McCarrol, S.A., Nemesh, J., Dermitzakis, E., Keinan, A., Montgomery, S.B., Pollack, S., Price, A.L., Soranzo, N., Bonnen, P.E., Gibbs, R.A., Gonzaga-Jauregui, C., Keinan, A., Price, A.L., Yu, F., Anttila, V., Brodeur, W., Daly, M.J., Leslie, S., McVean, G., Moutsianas, L., Nguyen, H., Schaffner, S.F., Zhang, Q., Ghori, M.J., McGinnis, R., McLaren, W., Pollack, S., Price, A.L., Schaffner, S.F., Takeuchi, F., Grossman, S.R., Shlyakhter, I., Hostetter, E.B., Sabeti, P.C., Adebamowo, C.A., Foster, M.W., Gordon, D.R., Licinio, J., Manca, M.C., Marshall, P.A., Matsuda, I., Ngare, D., Wang, V.O., Reddy, D., Rotimi, C.N., Royal, C.D., Sharp, R.R., Zeng, C., Brooks, L.D. & McEwen, J.E. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature,* 467(7311), pp. 52-8.

It, V., Barrientos, L., Lopez Gappa, J., Posik, D., Diaz, S., Golijow, C. & Giovambattista, G. (2010). Association of canine juvenile generalized demodicosis with the dog leukocyte antigen system. *Tissue Antigens,* 76(1), pp. 67-70.

Jia, X., Chen, S., Zhou, H., Li, D., Liu, W. & Yang, N. (2013). Copy number variations identified in the chicken using a 60K SNP BeadChip. *Anim Genet,* 44(3), pp. 276-84.

Karim, L., Takeda, H., Lin, L., Druet, T., Arias, J.A., Baurain, D., Cambisano, N., Davis, S.R., Farnir, F., Grisart, B., Harris, B.L., Keehan, M.D., Littlejohn, M.D., Spelman, R.J., Georges, M. & Coppieters, W. (2011). Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nat Genet.,* 43, pp. 405-413.

Karlsson, E.K., Baranowska, I., Wade, C.M., Salmon Hillbertz, N.H., Zody, M.C., Anderson, N., Biagi, T.M., Patterson, N., Pielberg, G.R., Kulbokas, E.J., 3rd, Comstock, K.E., Keller, E.T., Mesirov, J.P., von Euler, H., Kampe, O., Hedhammar, A., Lander, E.S., Andersson, G., Andersson, L. & Lindblad-Toh, K. (2007). Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet,* 39(11), pp. 1321-8.

Karlsson, E.K., Sigurdsson, S., Ivansson, E., Thomas, R., Elvers, I., Wright, J., Howald, C., Tonomura, N., Perloski, M., Swofford, R., Biagi, T., Fryc, S., Anderson, N., Courtay-Cahen, C., Youell, L., Ricketts, S.L., Mandlebaum, S., Rivera, P., von Euler, H., Kisseberth, W.C., London, C.A., Lander, E.S., Couto, G., Comstock, K., Starkey, M.P., Modiano, J.F., Breen, M. & Lindblad-Toh, K. (2013). Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. *Genome Biol,* 14(12), p. R132.

Kijas, J.M., Wales, R., Tornsten, A., Chardon, P., Moller, M. & Andersson, L. (1998). Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. *Genetics,* 150(3), pp. 1177-85.

Kim, K.S., Larsen, N., Short, T., Plastow, G. & Rothschild, M.F. (2000). A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits. *Mammalian Genome,* 11, pp. 131-135.

King, J.W.B. & Roberts, R.C. (1960). Carcass length in the bacon pig: Its association with vertebrae numbers and prediction from radiographs of the young pig. *Anim. Prod.,* 2, pp. 59–65.

Klungland, H., Vage, D.I., Gomez-Raya, L., Adalsteinsson, S. & Lien, S. (1995). The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mamm Genome,* 6(9), pp. 636-9.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J. & International Human Genome Sequencing, C. (2001). Initial sequencing and analysis of the human genome. *Nature,* 409(6822), pp. 860-921.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Genome Project Data Processing, S. (2009a). The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25(16), pp. 2078-9.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Subgroup, G.P.D.P. (2009b). The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25(16), pp. 2078-2079.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., 3rd, Zody, M.C., Mauceli, E., Xie, X., Breen, M., Wayne, R.K., Ostrander, E.A., Ponting, C.P., Galibert, F., Smith, D.R., DeJong, P.J., Kirkness, E., Alvarez, P., Biagi, T., Brockman, W., Butler, J., Chin, C.W., Cook, A., Cuff, J., Daly, M.J., DeCaprio, D., Gnerre, S., Grabherr, M., Kellis, M., Kleber, M., Bardeleben, C., Goodstadt, L., Heger, A., Hitte, C., Kim, L., Koepfli, K.P., Parker, H.G., Pollinger, J.P., Searle, S.M., Sutter, N.B., Thomas, R., Webber, C., Baldwin, J., Abebe, A., Abouelleil, A., Aftuck, L., Ait-Zahra, M., Aldredge, T., Allen, N., An, P., Anderson, S., Antoine, C., Arachchi, H., Aslam, A., Ayotte, L., Bachantsang, P., Barry, A., Bayul, T., Benamara, M., Berlin, A., Bessette, D., Blitshteyn, B., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Brown, A., Cahill, P., Calixte, N., Camarata, J., Cheshatsang, Y., Chu, J., Citroen, M., Collymore, A., Cooke, P., Dawoe, T., Daza, R., Decktor, K., DeGray, S., Dhargay, N., Dooley, K., Dooley, K., Dorje, P., Dorjee, K., Dorris, L., Duffey, N., Dupes, A., Egbiremolen, O., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Ferreira, P., Fisher, S., FitzGerald, M., Foley, K., Foley, C., Franke, A., Friedrich, D., Gage, D., Garber, M., Gearin, G., Giannoukos, G., Goode, T., Goyette, A., Graham, J., Grandbois, E., Gyaltsen, K., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Healy, C., Hegarty, R., Honan, T., Horn, A., Houde, N., Hughes, L., Hunnicutt, L., Husby, M., Jester, B., Jones, C., Kamat, A., Kanga, B., Kells, C., Khazanovich, D., Kieu, A.C., Kisner, P., Kumar, M., Lance, K., Landers, T., Lara, M., Lee, W., Leger, J.P., Lennon, N., Leuper, L., LeVine, S., Liu, J., Liu, X., Lokyitsang, Y., Lokyitsang, T., Lui, A., Macdonald, J., Major, J., Marabella, R., Maru, K., Matthews, C., McDonough, S., Mehta, T., Meldrim, J., Melnikov, A., Meneus, L., Mihalev, A., Mihova, T., Miller, K., Mittelman, R., Mlenga, V., Mulrain, L., Munson, G., Navidi, A., Naylor, J., Nguyen, T., Nguyen, N., Nguyen, C., Nguyen, T., Nicol, R., Norbu, N., Norbu, C., Novod, N., Nyima, T., Olandt, P., O'Neill, B., O'Neill, K., Osman, S., Oyono, L., Patti, C., Perrin, D., Phunkhang, P., Pierre, F., Priest, M., Rachupka, A., Raghuraman, S., Rameau, R., Ray, V., Raymond, C., Rege, F., Rise, C., Rogers, J., Rogov, P., Sahalie, J., Settipalli, S., Sharpe, T., Shea, T., Sheehan, M., Sherpa, N., Shi, J., Shih, D., Sloan, J., Smith, C., Sparrow, T., Stalker, J., Stange-Thomann, N., Stavropoulos, S., Stone, C., Stone, S., Sykes, S., Tchuinga, P., Tenzing, P., Tesfaye, S., Thoulutsang, D., Thoulutsang, Y., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Venkataraman, V., Vo, A., Wangchuk, T., Wangdi, T., Weiand, M., Wilkinson, J., Wilson, A., Yadav, S., Yang, S., Yang, X., Young, G., Yu, Q., Zainoun, J., Zembek, L., Zimmer, A. & Lander, E.S. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature,* 438(7069), pp. 803-19.

Magi, A., Benelli, M., Gozzini, A., Girolami, F., Torricelli, F. & Brandi, M.L. (2010). Bioinformatics for next generation sequencing data. *Genes (Basel),* 1(2), pp. 294-307.

Marklund, L., Moller, M.J., Sandberg, K. & Andersson, L. (1996). A missense mutation in the gene for melanocyte-stimulating hormone receptor (MC1R) is associated with the chestnut coat color in horses. *Mamm Genome,* 7(12), pp. 895-9.

Marklund, S., Kijas, J., Rodriguez-Martinez, H., Ronnstrand, L., Funa, K., Moller, M., Lange, D., Edfors-Lilja, I. & Andersson, L. (1998). Molecular basis for the dominant white phenotype in the domestic pig. *Genome Res,* 8(8), pp. 826-33.

McCarroll, S.A. & Altshuler, D.M. (2007). Copy-number variation and association studies of human disease. *Nat Genet,* 39(7 Suppl), pp. S37-42.

Merriman, B., D Team, I.T. & Rothberg, J.M. (2012). Progress in Ion Torrent semiconductor chip based sequencing. *ELECTROPHORESIS,* 33(23), pp. 3397-3417.

Mikawa, S., Morozumi, T., Shimanuki, S., Hayashi, T., Uenishi, H., Domukai, M., Okumura, N. & Awata, T. (2007). Fine mapping of a swine quantitative trait locus for number of vertebrae and analysis of an orphan nuclear receptor, germ cell nuclear factor (NR6A1). *Genome Res.,* 17, pp. 586-593.

Miller, W.H., Jr., Scott, D.W., Wellington, J.R. & Panic, R. (1993). Clinical efficacy of milbemycin oxime in the treatment of generalized demodicosis in adult dogs. *J Am Vet Med Assoc,* 203(10), pp. 1426-9.

Miller, W.H., Jr., Wellington, J.R. & Scott, D.W. (1992). Dermatologic disorders of Chinese Shar Peis: 58 cases (1981-1989). *J Am Vet Med Assoc,* 200(7), pp. 986-90.

Molin, A.M., Berglund, J., Webster, M.T. & Lindblad-Toh, K. (2014). Genome-wide copy number variant discovery in dogs using the CanineHD genotyping array. *BMC Genomics,* 15, p. 210.

Newton, J.M., Wilkie, A.L., He, L., Jordan, S.A., Metallinos, D.L., Holmes, N.G., Jackson, I.J. & Barsh, G.S. (2000). Melanocortin 1 receptor variation in the domestic dog. *Mamm Genome,* 11(1), pp. 24-30.

Olsson, M., Meadows, J.R., Truve, K., Rosengren Pielberg, G., Puppo, F., Mauceli, E., Quilez, J., Tonomura, N., Zanna, G., Docampo, M.J., Bassols, A., Avery, A.C., Karlsson, E.K., Thomas, A., Kastner, D.L., Bongcam-Rudloff, E., Webster, M.T., Sanchez, A., Hedhammar, A., Remmers, E.F., Andersson, L., Ferrer, L., Tintle, L. & Lindblad-Toh, K. (2011). A novel unstable duplication upstream of HAS2 predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs. *PLoS Genet,* 7(3), p. e1001332.

Online Mendelian Inheritance in Animals (OMIA) (09.06.2014.). http://omia.angis.org.au/home/ [13.07.2014.].

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B.M., Gray, J.W. & Albertson, D.G. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet,* 20(2), pp. 207-11.

Plant, J.D., Lund, E.M. & Yang, M. (2011). A case-control study of the risk factors for canine juvenile-onset generalized demodicosis in the USA. *Vet Dermatol,* 22(1), pp. 95-9.

Pryce, J.E., Hayes, B.J., Bolormaa, S. & Goddard, M.E. (2011). Polymorphic regions affecting human height also control stature in cattle. *Genetics,* 187, pp. 981-984.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. & Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet,* 81(3), pp. 559-75.

Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V. & Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics,* 28(18), pp. i333-i339.

Rubin, C.J., Megens, H.J., Martinez Barrio, A., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, O., Jern, P., Jorgensen, C.B., Archibald, A.L., Fredholm, M., Groenen, M.A. & Andersson, L. (2012). Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci U S A,* 109(48), pp. 19529-36.

Rubin, C.J., Zody, M.C., Eriksson, J., Meadows, J.R., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallbook, F., Besnier, F., Carlborg, O., Bed'hom, B., Tixier-Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K. & Andersson, L. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature,* 464(7288), pp. 587-91.

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S., Altshuler, D. & International, S.N.P.M.W.G. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* 409(6822), pp. 928-33.

Salmon Hillbertz, N.H., Isaksson, M., Karlsson, E.K., Hellmen, E., Pielberg, G.R., Savolainen, P., Wade, C.M., 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. *Nat Genet,* 39(11), pp. 1318-20.

Signer-Hasler, H., Flury, C., Haase, B., Burger, D., Simianer, H., Leeb, T. & Rieder, S. (2012). A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PLoS One,* 7(5), p. e37282.

Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider, G. & Ng, P.C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res,* 40(Web Server issue), pp. W452-7.

Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C., Tyler-Smith, C., Carter, N., Scherer, S.W., Tavare, S., Deloukas, P., Hurles, M.E. & Dermitzakis, E.T. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science,* 315(5813), pp. 848-53.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S. & DePristo, M.A. (2002). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In: *Current Protocols in Bioinformatics* John Wiley & Sons, Inc. Available from: http://dx.doi.org/10.1002/0471250953.bi1110s43.

Van Laere, A.S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A.L., Haley, C.S., Buys, N., Tally, M., Andersson, G., Georges, M. & Andersson, L. (2003). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature,* 425(6960), pp. 832-6.

Vaysse, A., Ratnakumar, A., Derrien, T., Axelsson, E., Rosengren Pielberg, G., Sigurdsson, S., Fall, T., Seppala, E.H., Hansen, M.S., Lawley, C.T., Karlsson, E.K., Consortium, L., Bannasch, D., Vila, C., Lohi, H., Galibert, F., Fredholm, M., Haggstrom, J., Hedhammar, A., Andre, C., Lindblad-Toh, K., Hitte, C. & Webster, M.T. (2011). Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet,* 7(10), p. e1002316.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. & Zhu, X. (2001). The sequence of the human genome. *Science,* 291(5507), pp. 1304-51.

Wang, K., Li, M. & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.,* 38, p. e164.

Weiling, F. (1991). Historical study: Johann Gregor Mendel 1822-1884. *Am J Med Genet,* 40(1), pp. 1-25; discussion 26.

Wilbe, M., Jokinen, P., Truve, K., Seppala, E.H., Karlsson, E.K., Biagi, T., Hughes, A., Bannasch, D., Andersson, G., Hansson-Hamlin, H., Lohi, H. & Lindblad-Toh, K. (2010). Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. *Nat Genet,* 42(3), pp. 250-4.

Yang, J., Meyer, M., Muller, A.K., Bohm, F., Grose, R., Dauwalder, T., Verrey, F., Kopf, M., Partanen, J., Bloch, W., Ornitz, D.M. & Werner, S. (2010). Fibroblast growth factor receptors 1 and 2 in keratinocytes control the epidermal barrier and cutaneous homeostasis. *J Cell Biol,* 188(6), pp. 935-52.

Yang, T.L., Guo, Y., Shen, H., Li, J., Glessner, J.T., Qiu, C., Deng, F.Y., Tian, Q., Yu, P., Liu, Y.Z., Liu, Y.J., Hakonarson, H., Grant, S.F. & Deng, H.W. (2013). Copy number variation on chromosome 10q26.3 for obesity identified by a genome-wide study. *J Clin Endocrinol Metab,* 98(1), pp. E191-5.

Zeitouni, B., Boeva, V., Janoueix-Lerosey, I., Loeillet, S., Legoix-ne, P., Nicolas, A., Delattre, O. & Barillot, E. (2010). SVDetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data. *Bioinformatics,* 26(15), pp. 1895-6.

Acknowledgements

This work was performed at the Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences (SLU). I would like to thank everyone I have had the opportunity to work with during these years, in particular I would like to acknowledge:

Göran Andersson, my main supervisor for choosing me as your PhD student. I couldn't have made it with anyone else. I would like to thank you for all interesting scientific discussions, guidance and giving me an opportunity to be part of so many interesting projects and introducing me to many collaborators. Especially, I would like to thank you for believing in me and providing me with an environment to learn and express my scientific views with freedom.

My co-supervisors:

Leif Andersson, for your precious scientific support, and guidance. I would like to thank you for all the scientific discussions and allowing me to take part in the most exciting projects. Your knowledge and enthusiasm have been a great inspiration for me during these years.

Carl Johan Rubin, for all the time you spent on discussing my projects and to help me improve my work. I would especially like to thank you for your patience and consistent guidance that has helped me to bridge between genetics and bioinformatics.

Kerstin Lindblad-Toh, for your scientific support and teaching me the clinical aspects of the disease. Especially for your valuable suggestions for writing and proofreading.

Erik Bongcam-Rudloff for introducing me to the world of bioinformatics and for your guidance and support throughout this time.

Álvaro Martinez Barrio, for your valuable suggestions and bioinformatic discussions. Thanks a lot for all your support.

Higher Education Commission of Pakistan for funding my PhD research.

Tomas Bergström for your scientific discussions and support especially with writing.

Kerstin Bergvall for sharing your data and helping me understand the clinical aspects of disease.

Gabriella Lindgren for providing me an opportunity to teach in your course Genome Analysis.

I would like to acknowledge all the people who I have worked with, in different groups, at our department SLU and at BMC (imbim:D11:3).

Finally, I would like to acknowledge and thank all my colleagues for their valuable discussions, especially **Jennifer** and **Matteo**. **Marcin** for your statistical guidance. **Agnese, Merina, Nima and Sangeet** (I really enjoyed your company in our trip to USA)

My very dear friend and colleague **Suzana Stjelja**, I will always cherish our discussions on GWAS and putting HINA on your hand ;) and Mehreen Ali.

All my **Pakistani colleagues**, **friends** and **families**, who have made this time so enjoyable.

And last but not the least my **parents** and **family**, especially my father and my husband **Jasim** for encouraging me to see the brighter side of life. Thank you for all your support, encouragement and motivation during all this time.