Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science

Identification and analysis of the dwarf mutation in domestic rabbits

Dou Hu

Examensarbete / Swedish University of Agricultural Sciences
Department of Animal Breeding and Genetics
444
Uppsala 2014

Master's Thesis, 30 hp

Erasmus Mundus Programme – European Master in Animal Breeding and Genetics

Identification and analysis of the dwarf mutation in domestic rabbits

Identifiering och karaktärisering av dvärg mutationen hos tamkaninen

Dou Hu

Supervisor:
Leif Andersson, SLU, Department of Animal Breeding and Genetics
Assistant Supervisors:
Carl-Johan Rubin, Uppsala University, Department of Medical Biochemistry and Microbiology.
Miguel Carneiro, University of Porto, CIBIO.
Examiner:
Erling Strandberg, SLU, Department of Animal Breeding and Genetics

Credits: 30 HEC
Course title: Degree project in Animal Science
Course code: EX0556
Programme: Erasmus Mundus programme
– European Master in Animal Breeding and Genetics
Level: Advanced, A2E

Place of publication: Uppsala
Year of publication: 2014
Cover picture: Miguel Carneiro
Name of series: Examensarbete / Swedish University of Agricultural Sciences
Department of Animal Breeding and Genetics, 444
On-line publication: http://epsilon.slu.se

Key words: dwarf mutation; *HMGA2*; domestic rabbis; embryos

European Master in Animal Breeding and Genetics

IDENTIFICATION AND ANALYSIS OF THE DWARF MUTATION IN DOMESTIC RABBITS

DOU HU
900507-P122

THESIS ANIMAL BREEDING AND GENETICS (EX0556)
MAY 2014

Department of Animal Breeding and Genetics

SUPERVISORS
Supervisor 1 Leif Andersson
Supervisor 2 Miguel Carneiro
Supervisor 3 Carl-Johan Rubin

Education and Culture
Erasmus Mundus

CONTENTS

ABSTRACT

The Dwarf mutation in domestic rabbits is caused by a semi-lethal autosomal recessive single gene on chromosome 4. In the heterozygote state the mutation causes dwarfism and in the homozygote state the mutation gives an extreme dwarf phenotype (peanut) not compatible with life for more than a few days post-partum. In this study whole genome re-sequencing was used for mapping and identification of the causal dwarf mutation. A 12.1Kb deletion spanning between 44,709,089 bp to 44,721,236 bp on chromosome 4, taking out three exons of the *High-mobility group AT-hook 2 (HMGA2)* gene was identified as the causal mutation. Genotyping of the *HMGA2* deletion in more dwarf, peanut and wild-type rabbits further supported the causality of the deletion. We show that the dwarf phenotype arises from the inactivation of *HMGA2*, a member of the high mobility group AT-hook family which function as an architectural factor in growth regulation during embryonic development. Expression profile analysis of different parts of wild-type embryos revealed high expression levels of *HMGA2* in the early developmental stages from day 9.5 to day 18 and a dramatic decrease from day 21. Embryos at day 15.5 displayed the highest expression level of *HMGA2*. We observed low expression levels in brain tissue and relative expression levels in different body parts in embryos (day 21 and 24) as well as in newborn rabbits could be ordered in the following way: body > head skin > skull > brain.

Key words: dwarf mutation; *HMGA2*; domestic rabbits; embryos

1. INTRODUCTION

1.1 The Dwarf phenotype

Growth is the result of complex interactions between multiple environmental and genetic factors. Dwarfism is present in many species, such as human, mice, chicken and rabbit (Ruyter-Spira *et al.* 1998). The pioneering work on dwarfism in rabbits can be traced to 1934 when Greene, Hu, and Brown reported that the dwarf phenotype in domestic rabbits shows an autosomal semi-lethal inheritance (Greene *et al.* 1934). The heterozygous carriers of the dwarf allele typically are approximately two-thirds the size of normal litter mates when they are born; have compact and rounded bodies, short noses, a disproportionate larger head when compared to rest of the body, and small ears (Figure 1A). Animals that are homozygous for the dwarf allele show a severely reduced growth rate, with approximately one-third the size of their normal litter mates and less than half the birth weight of their normal sibs. They always display a very characteristic body conformation evident at the time of birth that makes them easily to identify (Figure 1B). Homozygous mutants have rounded posterior calvarium, a dished-out snout, extremely small ears and an abnormal configurative head (Greene 1940). They are viable up to the time of birth but they die within the first few days. In has been reported that the secretory functions of the pituitary is completely inhibited which underlie the peanut phenotype (Greene 1940).

There have been many studies examining the genetic architectures of mammalian dwarf phenotypes. For instance, studies of the 12q14 micro-deletion syndrome revealed the roles of *HMGA2* in regulating linear growth in humans. Affected patients with the 12q14 micro-deletion syndrome show a short stature (Tay *et al.* 2009). The Crooked Neck Dwarf mutation in chickens is an autosomal recessive lethal mutation caused by one single gene conferring skeletal muscle dysgenesis including muscle dysfunction, a failure to maintain embryonic skeletal tendons and muscles, and eventually the degenerative loss of all skeletal muscles (Airey *et al.* 1993). Another spontaneous autosomal recessive mutation, found in White Leghorn chickens, results in reduced 30% reduction in body weight in adults and overall decreased body size except the head (Ruyter-Spira *et al.* 1998). The small size in pygmy mice cannot be explained by aberrations in the growth hormone-insulin-like growth factor endocrine pathway (Zhou *et al.* 1995). In 1995, Zhou demonstrated that the mice pygmy phenotype was caused by genetic variation in *high-mobility group protein I-C (HMGI-C)* gene which has been shown to be involved in the regulation of cell proliferation (Zhou *et al.* 1995).

Previous studies have confirmed that the dwarf phenotype of domestic rabbits is governed by the effects of a single autosomal recessive dwarf gene (d), located on chromosome 4. Previous linkage data revealed that the genes for agouti hair pattern and the dwarf phenotype are located on the same chromosome, with a reported crossover percentage of 12-15% (Castle & Sawin 1941). However, studies on the specific genomic region harbouring the dwarf mutation and the casual mutation underlying the dwarf phenotypes were unknown when this study was initiated.

Figure 1: A: A New Zealand rabbit and a dwarf rabbit. Left is the normal rabbit; right is the dwarf rabbit with compact and rounded body, small ears, short nose, rounded posterior calvarium and a disproportionate large head comparing with rest of the body. B: A homozygous peanut (dd) rabbit with extremely large head and tiny ears, misshapen small front legs and deformed back legs.

1.2 Whole genome re-sequencing

Sanger capillary based sequencing was been the most commonly used DNA sequencing method for nearly 30 years since 1977 (Schuster 2007). However, since several new sequencing instruments became available in the first decade of the 21st century, the so-called 'next generation' or 'massively parallel' sequencing methodologies are becoming widely used and already improving the field (Mardis 2008). The next-generation sequencing instrumentation are capable to re-sequence genomes in a fraction of the time, effort and expense than was possible using capillary sequencers (Hillier *et al.* 2008).

Next-generation sequencing methods overcome the limitation of traditional Sanger sequencing by either attaching the DNA molecules on surfaces or creating micro-reactors to produce millions of sequencing reactions (Reis-Filho 2009). Nowadays, there are three main technologies commercially used. Roche (454) GS FLX sequencer was introduced in 2004 and worked on the concept of pyro-sequencing to produce an average read length of ~250 bp per sample with a combined throughput of ~100MB of sequence data per 7h run. The Illumina Genome Analyzer was introduced in 2006 and worked on the principle of sequencing by synthesis to produce ~32-40 bp sequence reads from tens of millions of surface-amplified DNA fragments. Applied Biosystems SOLiD sequencer was released in 2007 and utilized a special sequencing process catalysed by DNA ligase. Each SOLiD run produces 3-4 Gb sequencing data with an average read length of ~25-35 bp in ~5 days (Mardis 2008).

The Human Genome Sequencing Consortium generated 3Gb at the cost of approximately $3 billion and took 13 years by traditional sequencing method. Current next-generation sequencing technology could obtain 10-fold coverage of the human genome (30GB DNA sequence) in a single run for less than $15,000 to $20,000 (Reis-Filho 2009). In 2014, the cost has already decreased to $2,000. Whole genome re-sequencing technology has improved rapidly and is now two orders of magnitude faster and more cost effective than the technologies originally used for the sequencing of the human genome and it is expected to reduce cost to sequence an entire human genome for $1,000 in the near future (Meuwissen & Goddard 2010).

In recent years next-generation sequencing has demonstrated its enormous potential for functional genomics studies in animals (Schuster 2007). Whole genome sequencing or re-sequencing has been performed in many livestock species, such as pigs, chickens and cattle. It also brings new opportunities and enormous possibilities for identification of causative mutation for complex traits in domestic animals.

1.3 *HMGA2*

HMGA2, also known as HMGI-C, a member of the high mobility group AT-hook family of non-histone chromatin proteins, is an architectural transcription factor (Lee & Dutta 2007). The HMGA proteins contain three DNA-binding domains, termed AT hook domains, which interact with AT- rich stretches in the narrow minor groove of DNA and an acidic C-terminal tail (Brants *et al.* 2004). It has been proven that *HMGA2* protein plays a significant role in growth regulation during embryonic development (Zhou *et al.* 1995; Lee & Dutta 2007).

In humans, a chromosomal rearrangements at chr12q13-15 involving the *HMGA2* gene (Alyaqoub *et al.* 2012) and the accompanying overexpression of the HMGA2 protein result in benign mesenchymal tumors such as uterine leiomyoma, endometrial polyps, lipoma and pulmonary chondroid hamartoma (Lee & Dutta 2007). Except its critical role in benign mesenchymal

tumorigenesis, there are a number of appealing observations with regard to *HMGA2* in malignant epithelia tumors over the past decades (Morishita *et al.* 2013). Furthermore, overexpression of *HMGA2* is tumorgenic in differentiated tissues which is correlated to the occurrence of colon cancer, breast cancer, lung cancer, myeloid neoplasia and oral carcinomas (Miyazawa *et al.* 2004; Morishita *et al.* 2013). *HMGA2* can be utilized as a biomarker of melanoma progression and prognosis as well (Raskin *et al.* 2013). Regulation of *HMGA2* expression during prenatal development might be one determining factor for human stature (Tay *et al.* 2009).

HMGA2, as an effector of TGF-β that causes Epithelial-mesenchymal transition (EMT), is expressed at high level by transformed cells or tumors of mesenchymal and epithelial origin even though *HMGA2* is expressed during embryogenesis and becomes silent in the mature tissues (Sgarra *et al.* 2004; Thuault *et al.* 2006). EMT appears during embryogenesis, carcinoma invasiveness, and metastasis and can be elicited by transforming growth factor-β (TGF-β) signalling via intracellular Smad transducers and can convert polarized epithelial cells to motile mesenchymal cells (Thuault *et al.* 2006).

In mice, the pygmy mutation is unique among mutations leading to dwarfism because its phenotype cannot be explained by aberrations in the growth hormone-insulin-like growth factor endocrine pathway but is caused by the inactivation of *HMGA2* (Zhou *et al.* 1995; Brants *et al.* 2004). *HMGA2* gene is encoded by the pygmy locus on mouse chromosome 10 and its expression is high during early development from 10.5 d.p.c to 14.5 d.p.c but barely detectable in adult tissues (Zhou *et al.* 1996). Previous studies demonstrated that most tissues in pygmy mouse were 40% to 50% smaller than in wild-type mouse tissues except for the brain which stays the normal size (Benson & Chada 1994). The insulin-like growth factor mRNA-binding proteins (IMP) has an important role for regulating growth, and is composed of three family members IMP1, IMP2 and IMP3. Only IMP2 is down-regulated in pygmy mutant embryos lacking *HMGA2* expression. The *HMGA2-IMP2* axis has been shown to play a key role in regulating satellite cell activation and therefore skeletal muscle development (Li *et al.* 2012). Disturbing the axis may results in a skeletal muscle dysgenesis involving muscle dysfunction and a failure to maintain embryonic skeletal tendons and muscles.

1.4 Aims of the study

- To identify the genomic region harbouring the dwarf mutation and to identify the causal mutation underlying the dwarf and peanut phenotypes by whole genome re-sequencing of normal and dwarf rabbits
- Genotyping of a large number of samples representing the different dwarf phenotypes (dwarfs and peanuts) and normal rabbits to confirm the causal dwarf mutation
- Determination of the expression profiles of *HMGA2* gene in wild-type rabbit embryos at different developmental stages and different body sites

Table 1: Selected papers studying mutations in HMGA2 in mammals

Time	References	Event and breakpoint about dwarf mutation or *HMGA2*
1934	Greene *et al.*	First described an incomplete lethal recessive mutation and designated it dwarf in domestic rabbits
1940	Harry *et al.*	Extensive investigation of the lethal dwarf mutation on physical appearance and histological analysis in Polish rabbits
1941	W. E. Castle *et al.*	Described a linkage system involving agouti (A) gene and dwarf (d) gene; the crossover percentage between agouti and dwarf was around 12%-15% in rabbits
1995	Zhou *et al.*	Absence of *HMGA2* expression in mice resulted in the pygmy phenotype; high expression of *HMGA2* between 10.5 d.p.c to 14.5 d.p.c in wide type; most tissues in pygmy mice were 40%-50% smaller than wild-type tissues except for brain tissue;
1996	Zhou *et al.*	Described *HMGA2* genomic structure in mice; RNA-seq was used to analysis temporal and tissue-specific expression of *HMGA2* during murine development
2002	Chieffi *et al.*	Disruption of *HMGA2* gene resulted in a block of spermatogenesis and a pygmy phenotype with a drastic reduction in fat tissue and was associated with a longer cell cycle of embryonic fibroblasts
2004	Jan R. Brants *et al.*	Only IMP2, not its family member IMP1 and IMP3 was down-regulated in mutant E12.5
2007	Y. Lee *et al.*	*HMGA2* was depressed upon inhibition of *let-7* in cells with high levels of the miRNA; described the genomic structure and location of *HMGA2* in human
2007	C. Mayr *et al.*	*Let-7* miRNA acted as a tumor-suppressor gene; indicated that a major mechanism of oncogenic *HMGA2* translocations associated with various human tumors was the loss of *let-7* repression
2012	Z. Li *et al.*	*HMGA2*-IMP2 axis functions as a key regulator of satellite cell activation and therefore skeletal muscle development
2013	Morishita A *et al.*	*HMGA2* is an architectural transcription factor predominantly expressed in the mesenchyme before its differentiation and a regulator of mesenchymal proliferation and differentiation; the *HMGA2* null reveals a pygmy phenotype due to the decreased number of mesenchymal cells in mice

2. MATERIALS AND METHODS

2.1 Animals

Rabbits that are homozygous for the wild-type allele at the dwarf locus have the genotype DD. Rabbits that are heterozygous Dd show the dwarf phenotype while dd homozygotes are called "Peanut".

In this study, DNA that had been extracted from blood samples was collected from 10 dwarf rabbits (Dd) and 10 peanut rabbits (dd). These DNA samples were separately mixed as two DNA pools and submitted for whole genome re-sequencing to identify candidate mutations for the dwarf phenotype. Following the detection of the candidate causal *HMGA2* deletion another 20 Netherland dwarf rabbits (14 females and 6 males), 14 peanut samples (8 females and 6 males), and 20 wild type rabbits were further genotyped to confirm the causality of mutation. The wild type rabbits were comprised of 3 wild rabbits from Porto Santo island, 3 domestic Champagne rabbits, 3 wild Toledo rabbits, 3 French wild Caumont rabbits, 4 French wild Villemolaque rabbits and 4 Calzada rabbits (Appendix Table 3).

11 embryos from wild type rabbits at different developmental stages were collected for expression profile analysis, including 3 embryos at day 9.5, two embryos at day 12 and one embryo at day 15.5, 16, 18, 21, 24 and new born respectively.

2.2 Mapping and identification of the causal mutation by Whole Genome Re-sequencing

The two DNA pools described above were sequenced as paired-end libraries to 30X coverage using a HiSeq2 instrument (Illumina). All reads were aligned to the rabbit reference genome assembly (OryCun2) using the software BWA-mem (Li & Durbin 2009). Bioinformatic analysis was used to identify all sequence variants, such as structural changes, insertion/deletions and SNPs, which were unique to the dwarf haplotype (d) and not found in any wild-type rabbits (DD). Samtools was used to generate a pileup file and then the PoPoolation package was utilized to calculate the frequency of candidate mutations for the dwarf phenotype in rabbits (Kofler *et al.* 2011).

2.3 Determination of breakpoint of deletion

Primers 1 and 5 were designed using Primer 3.0 (http://bioinfo.ut.ee/primer3-0.4.0/primer3/) for amplifying a fragment across the deleted region in the dwarf homozygotes to precisely define the deletion breakpoints (Primer design is indicated in Figure 2). PCR mix 1 (total reaction volume of 20µl) contained approximately 40ng DNA, 0.8µl primer mix (0.2µM and 0.4µl each), 2µl 10*Buffer, 1.6µl Mg^{2+} (2.0mM), 0.2µl dNTP (0.2mM), 0.25µl Taq polymerase and 14.15µl water. PCR amplification was carried out under the following thermo cycling profile: an initial

denaturation at 95°C for 9 min; 45 cycles of 94°C for 30s, 60°C for 1 min; final extension at 60°C for 10 min.

2.4 Genotyping

2.4.1 PCR amplification

PCR assays were used to genotype individual rabbit based on presence/absence of the *HMGA2* deletion breakpoints to confirm a complete concordance between the presence of this mutation and the dwarf and peanut phenotypes. Two pairs of primers were designed for genotyping based on the rabbit reference genome assembly (OryCun2) using Primer 3.0 (http://bioinfo.ut.ee/primer3-0.4.0/primer3/). These two primer pairs (P1+P2 and P1+P3) were used to generate amplicons for sequencing across the deletion breakpoint and this amplicons was also used as a diagnostic test which could be used to genotype all the dwarf samples, peanuts and wild-type rabbits.

PCR mix 2 contained 3µl DNA sample (~50ng), 0.4µl primer mix, 2µl 10*Buffer, 1.2µl Mg^{2+}, 0.2µl dNTP, 0.25µl Taq polymerase and 12.95µl water. PCR mix 3 contained 2µl sample from PCR mixes 2 products, 0.8µl primer mix, 2µl 10*Buffer, 1.2µl Mg^{2+}, 0.2µl dNTP, 0.25µl Taq and 13.55µl water. For the diagnostic test, the PCR mix 1 with P1+P3 was utilized to genotype peanuts and wild-type rabbits. Due to the low concentration of dwarf samples, nested PCR was used. PCR mix 2 with P1+P2 was used for the first round and PCR mix 3 with P1+P2 was used for the second round PCR to genotype dwarf phenotypes. The same PCR thermo cycling profile as used for the determination of breakpoints was also used here.

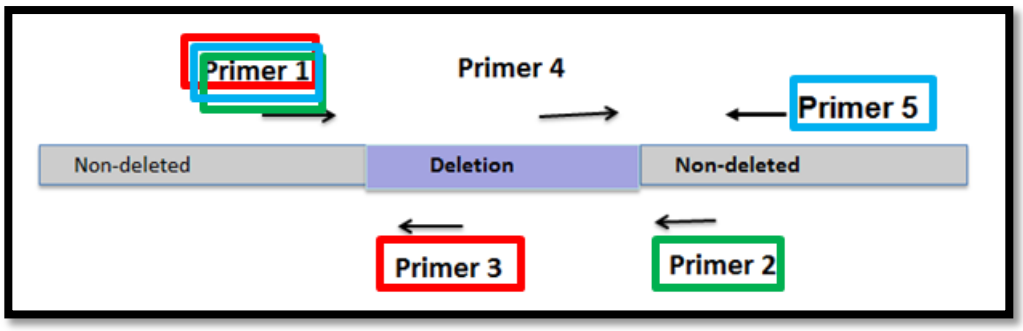

Figure 2: Primer design for genotyping and breakpoint analysis. The same colour box means one pair of primer. P1+P5 were used for determining the breakpoint; P1+P3 were used for genotyping peanuts and wild-type rabbits; P1+P2 were used for genotyping dwarf phenotypes.

2.4.2 Gel electrophoresis

3µl PCR amplification product mixed with loading dye was used for gel electrophoresis. 100bp molecular ruler was utilized in the study. P1+P2 produced a 293bp amplicon; P1+P3 produced a

244bp amplicon; and P1+P5 produced a 530bp amplicon.

2.5 Expression profile analysis

2.5.1 RNA extraction

Different body parts of embryos at different developmental stages were collected for RNA extraction as explained in Table 2.

Table 2: Embryo samples at different developmental stages used for gene expression analysis

	New born	Day 24	Day 21	Day 18	Day 16	Day 15.5	Day 12		Day 9.5		
							1	2	1	2	3
Body tissue	Dorsal Muscle[a]	Dorsal Muscle	Dorsal Muscle	Dorsal Muscle	Half body[b]	Half body	Whole embryo		Whole embryo		
Skull tissue	Skull	Skull		--	--	--	--		--		
Head skin	h-skin[c]	h-skin	Whole head[d]	--	--	--	--		--		
Brain tissue	Brain	Brain		--	--	--	--		--		

[a]: dorsal muscle tissues were cut from individuals

[b]: due to the limitation of body size, half of the body was used

[c]: head skin tissues were dissected from individuals

[d]: because it is difficult to dissect different parts of the head, the whole head was used

RNeasy® Mini Kit for animal tissues was used for purification of demanded RNA in this study. Details about steps in RNA purification can be seen in RNeasy website (http://www.genome.duke.edu/cores/microarray/services/rna-qc/documents/RNeasy_Mini_Handbook.pdf).

2.5.2 Reverse transcription and Quantitative PCR

Reverse transcription solution (total reaction volume of 20µl) contained Oligo$(T)_{18}$ primers (0.25µl), 10 mM dNTP (1µl), 5xRT buffer (4µl), M-MLV enzyme (1µl, Invitrogen^{TM}), template RNA (1µg) and nuclease-free water. Reverse transcription used the following thermo cycling profile: 25^{o}C for 10 min, 50^{o}C for 30 min and 85^{o}C for 5 min.

Template cDNA obtained from the Reverse transcription was utilized for quantitative PCR (qPCR). GAPDH was used as endogenous control. New *HMGA2* reverse and forward primers and HYB oligo TaqMan probe were designed for qPCR (Figure 3). Primer 3.0 was used for designing probe and one pair of *HMGA2* primers. Primers *HMGA2* Ex-F and *HMGA2* Ex-R were designed for crossing the region from exon 4 to exon 6 of *HMGA2*. A TaqMan Probe targeting exon 5 of *HMGA2* gene was designed. Details about primers and probe sequences are given in Appendix Table 3.

The PCR mix contained 5µl Taq universal PCR master mix (2X), 0.2µl reverse primer, 0.2µl forward primer, 0.01µl TaqMan probe, 3.59µl nuclease-free water and1µl cDNA sample. Applied Biosystems 7900HT real-time PCR system was used in this study.

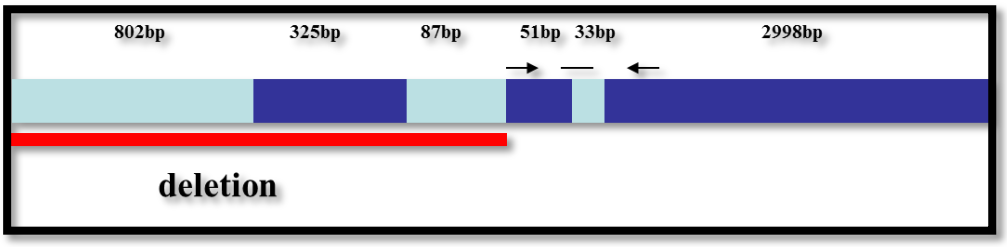

Figure 3: Primer design and probe design for expression profile analysis. Arrows means primer and black line means probe. Different colours of rectangles mean exons of HMGA2 gene. Red line means the candidate dwarf mutation in this study.

3. RESULTS

3.1 Mapping and identification of candidate causal mutations for rabbit dwarf phenotype

We used whole genome re-sequencing to identify the causal region responsible for the dwarf phenotype. BWA-mem was used to map the reads onto the rabbit reference genome, then, a pileup file was generated by samtools to calculate the F_{st} values between peanuts and dwarfs for the entire genome by PoPoolation package (Li & Durbin 2009; Kofler *et al.* 2011). A previous study suggested that the Dwarf locus is linked to Agouti on chromosome 4 (Castle & Sawin 1941), and the initial efforts were therefore directed to scans of this chromosome. After a series of bioinformatics analysis mentioned above, the causal region of dwarf mutation was revealed as a deletion in chromosome 4 in rabbits (Figure 4).

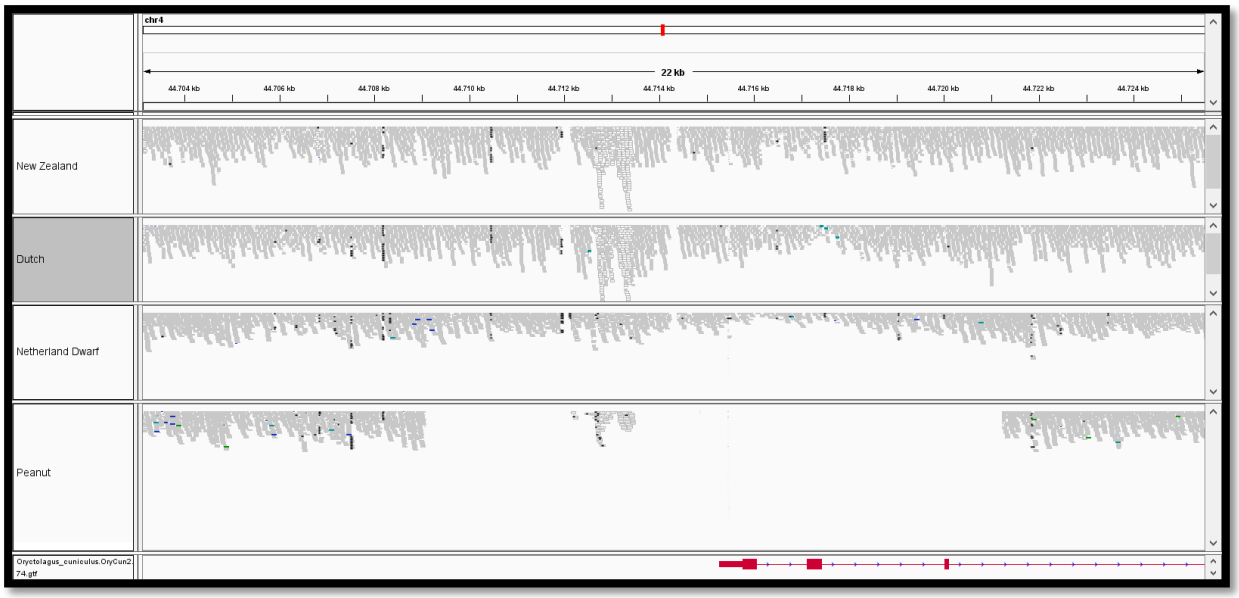

Figure 4: IGV (Thorvaldsdóttir et al. 2013) view of comparisons among New Zealand rabbit (WT/WT), Dutch rabbit (WT/WT), Netherland Dwarf (WT/Del) and Peanut (Del/Del). An obvious homozygous deletion was found in the peanut sample for the region from 44,709kb to 44,721kb on chromosome 4; Netherland Dwarf rabbits showed half expression level compared with wild type rabbit in the same region.

When analysing depth of read coverage for the two sequenced pools a deletion overlapping parts of the coding sequence of the High Motility Group Antigen 2 (*HMGA2*) gene was identified (Figure 4). Approximately 12 kb of the wild type sequence had no high confidence mapped reads in the Peanut pool homozygous for the deletion (Del/Del). The Dwarf pool had about half the expected read depth of coverage over this region, consistent with their known heterozygous genotypes expected for a causal mutation (Del/WT). Meanwhile, homozygous wild type rabbits of New Zealand and Dutch rabbits (WT/WT) revealed normal depth of coverage, approximately twice the depth of mapped reads compared with Netherland Dwarf rabbits in the same region. The deletion region is visible round 44,709 kb to 44,721 kb (Figure 4). The identified deletion overlaps the first three exons of the *HMGA2* gene making it very likely that this deletion is responsible for the two phenotypes associated with the rabbit dwarf locus in domestic rabbits. Further bioinformatics analysis determined that the deletion spanned between bases 44,709,089 bp to 44,721,236 bp on chromosome 4.

3.2 Validation of breakpoints by PCR and sequencing

All the dwarf samples and Peanuts were amplified by standard PCR with primers P1+P5 to determine the breakpoint.

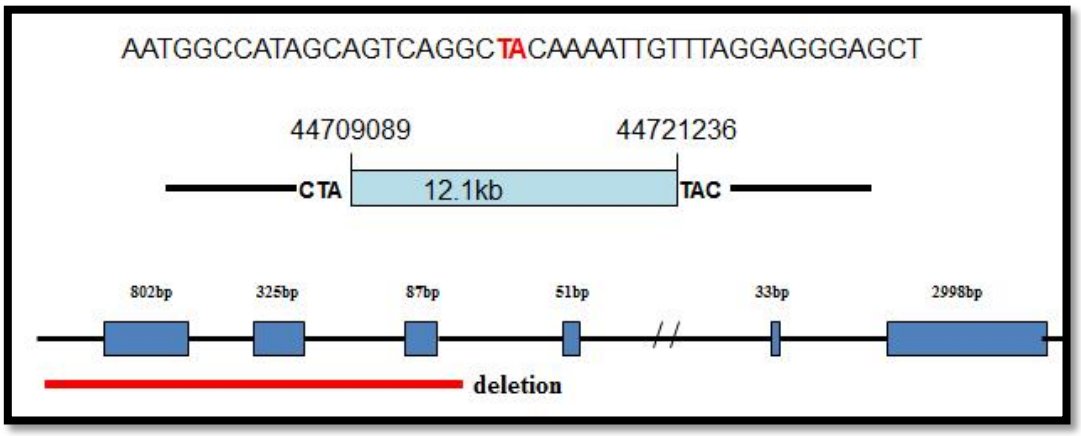

Figure 5: Sequencing result of selected samples. All three sequencing results showed the same sequence at the deletion breakpoint (red colour letters) at 44,709,089bp and 44,721,236bp (middle). The bottom figure shows a schematic illustration indicating that the deletion removes exon 1-3 of HMGA2 gene; dark blue rectangles mean exons.

PCR products from Peanut 13, Peanut 14 and Dwarf 50 (details in Table 3 in Appendix) were used for sequencing (Figure 5). The upper nucleotide sequence shows the consensus breakpoint derived from sequencing three samples and the middle schematic figure displays the sequence including the deletion part. The red nucleotides show the breakpoint of the deletion region as we determined in the bioinformatics analysis. The deletion starts at 6.1 kb of upstream of exon 1 and ends at 1.1kb downstream of exon 3 (Figure 5).

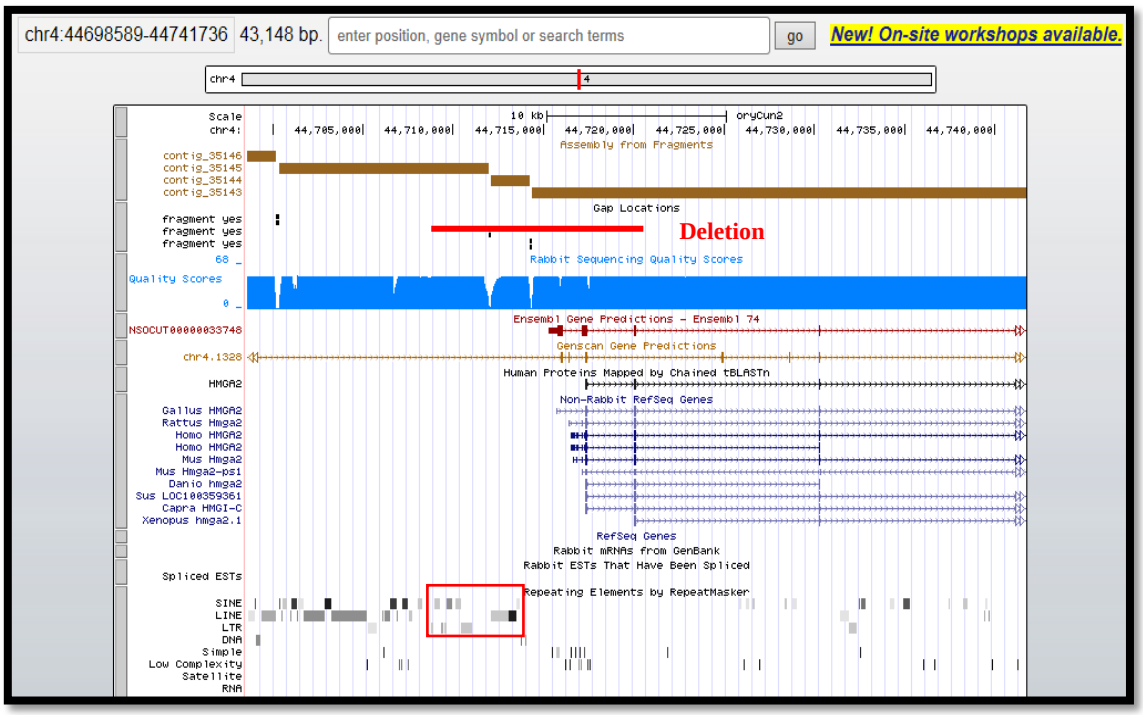

Figure 6: Genomic structure at the deletion site illustrated using the UCSC genome browser.

Red box shows that four SINEs and one LINE are detected close to the start site of the deletion. Deletion region was shown as the red line in the figure.

Genomic structure near the deletion was analysed by manual inspection using the UCSC genome browser. At the start site of the deletion, four SINEs and one LINE were found (Figure 6) which could be one explanation why the deletion event took place at this position.

3.3 Genotyping results

All 14 peanuts samples, 20 dwarf samples and 20 wild type samples were genotyped. When all the samples were subjected to PCR with primers P1+P2 and P1+P3 separately, the results were as follows: all peanuts samples showed only the 293bp amplicon; all dwarf samples showed both the 293bp and the 244bp amplicons; all wild type samples showed only the 244bp amplicon.

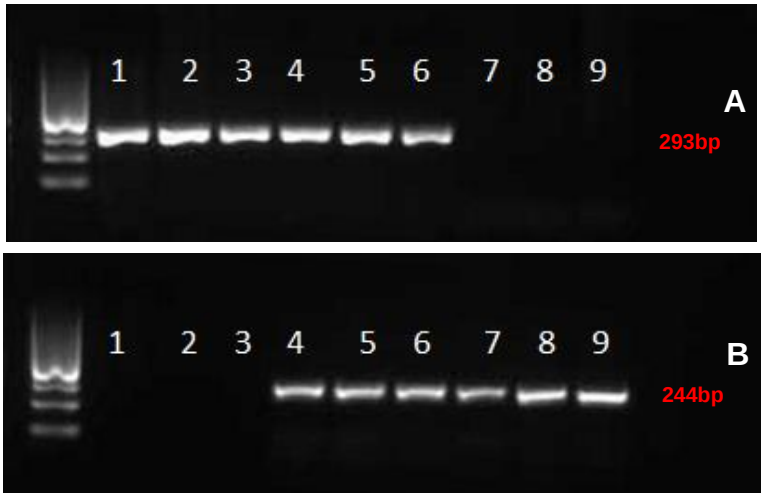

Figure 7: Genotyping results. Sample 1-3: PNT9, PNT10, PNT 11(peanut); Sample 4-6: NDA39, NDA41, NDA50 (dwarf); Sample 7-9: PST7, PST10, PST16 (wild type). A) PCR amplicons of Peanuts and dwarfs with P1+P2 were 293bp; no amplicons of wild type rabbits; B) PCR amplicons of dwarfs and wild type rabbits with P1+P3 were 244bp; no amplicons of Peanuts.

Only three samples which were selected from each phenotype group were utilized to show the genotyping results here in Figure 7. Based on the results we could conclude that all Peanuts were homozygous for deletion, dwarfs were heterozygous for deletion and wild-type rabbits were homozygous without deletion. All 14 peanuts samples, 20 dwarf samples and 20 wild type samples were perfectly consistent with the phenotypes. Thus, our results are regards the presence of the 12.1 kb deletion on chromosome 4 is consistent with our hypothesis that it is the causal mutation for the dwarf and peanut phenotypes in domestic rabbits.

3.4 Expression profile analysis

3.4.1 Confirmation of RNA quality

RNA was extracted from different body parts of 11 wild-type embryos as mentioned above (Table 2). All the extracted RNAs were run on gel to check the quality before expression profile analysis. Two apparent bands corresponding to 28s and 18s rRNA could be observed which demonstrated that extracted RNA was of good quality for all the samples (Figure 8).

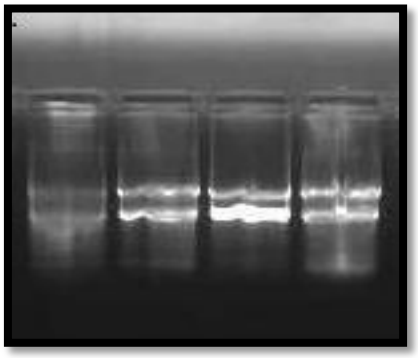

Figure 8: Gel electrophoresis results of RNA of body tissues in new born, 24, 21 and 18 days. There were two bands shown 28s and 18s rRNA.

Only 2µl RNA from each sample was used for gel electrophoresis, which might be the explanation for the inconspicuous fluorescence in one lane in Figure 8. However, the obvious two bands in all the samples demonstrate that all the extracted RNAs were of good quality.

3.4.2 Expression profile analysis of body tissues

In order to check where we could observe the highest expression level of *HMGA2* gene in domestic rabbits, body tissues were collected from different embryonic stages to do expression profile analysis. High expression levels of *HMGA2* could be observed during early development from day 9.5 to day 18. Embryos at day 15.5 displayed the highest expression level with a standard variance less than 0.2, which indicates that the result was reliable. The *HMGA2* expression level dramatically decreased from day 21 (Figure 9), which is consistent with previous studies in mice where *HMGA2* is barely detectable in adult tissues.

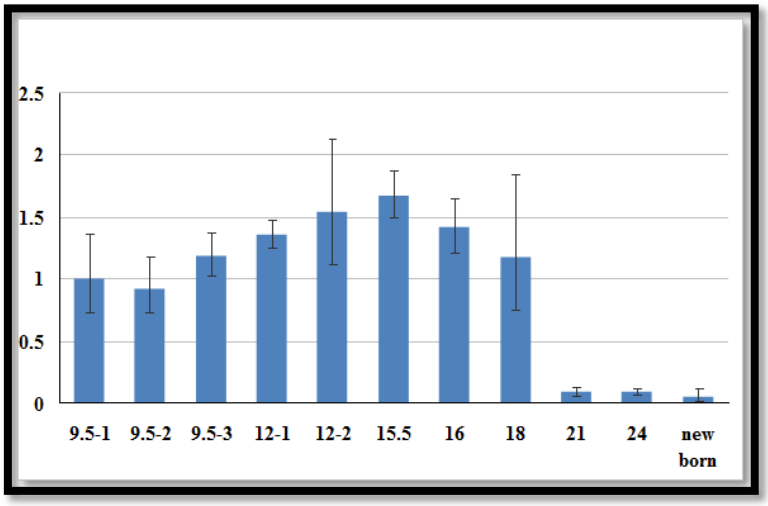

Figure 9: Expression profile analysis of body tissues in different embryo days. HMGA2 gene has the highest expression level at day 15.5. Its expression level dramatically decreases from day 21.

3.4.3 Expression profile analysis of different body parts

Earlier studies have demonstrated that most tissues in pygmy mouse were 40% to 50% smaller than wild-type mouse tissues except for the brain tissue which maintain its normal size (Zhou *et al.* 1995). We decided to explore whether the *HMGA2* expression pattern in rabbit brain during development differed from that in other for tissues. Skull tissue, head skin and brain tissue of embryos at day 24 and new born were collected for this purpose. In consideration of the small head size of embryos at day 21, half the head was collected to generate a comparison of *HMGA2* expression level among different brain tissues at various days.

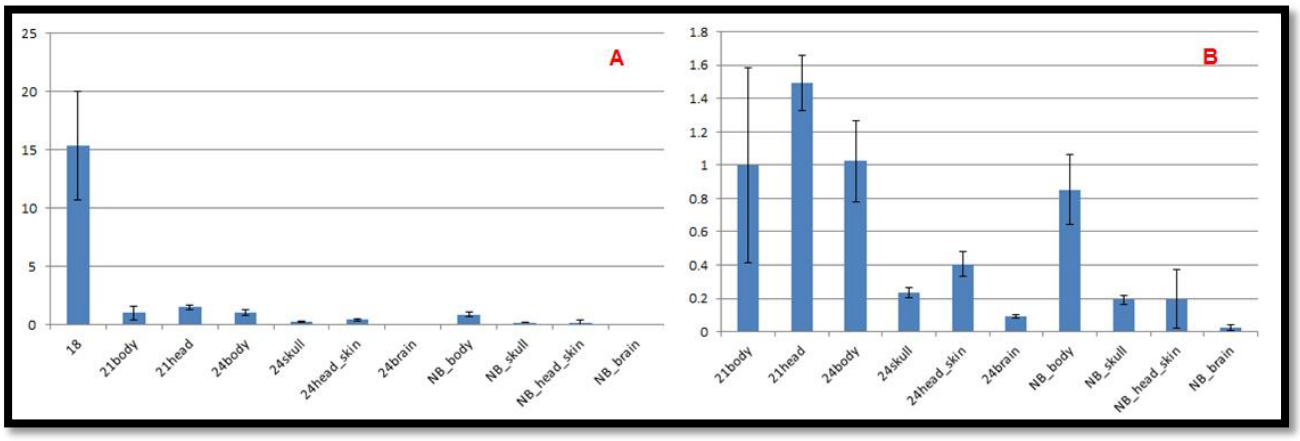

Figure 10: Expression profile of brain tissues at different developmental stages. HMGA2 gene has low expression level in brain tissues. A is with day 18 embryo as a positive control while B is without day 18 embryo. The expression level of HMGA2 in the different parts of brains at day 21, 24 and new born (NB) embryos: body > head skin > skull > brain.

Based on the results shown in Figure 10A, body tissue at day 18 was chosen as a positive control to reveal that *HMGA2* expression in both body tissue and brain tissues were dramatically decreased on day 21. To compare the expression level among body tissue and different brain tissues, body tissue at day 18 was extracted to impel the difference between diversified tissues becoming obvious. In figure 10B, the order of the expression level of *HMGA2* in different parts of embryos at day 21, 24 and new embryos is: body > head skin > skull > brain. Because of a pipetting error for body tissue at day 21, the expression level of *HMGA2* in head was higher than those in body. Repeated real-time PCR was done which illustrated that body tissue at day 21 had a higher expression level than head tissue with lower standard variation.

4. DISCUSSION

Dwarfism is a common trait in domestic rabbits in particular among pet rabbits. The first studies on the dwarf phenotype in domestic rabbits was published already 1934 (Greene, Hu and Brown 1934) and in 1941 (Castle & Sawin 1941). These studies showed that the dwarf phenotype is caused by a dwarf gene (d) which was a semi-lethal autosomal recessive single gene. However, the genomic region harbouring the dwarf mutation and the causal mutation underlying different phenotypes has been unclear until the present study.

Here we report about a successful application of next-generation sequencing technology for identifying the causal mutation responsible for dwarf phenotype in rabbits. Whole genome re-sequencing and bioinformatic analysis revealed a 12.1Kb deletion between 44,709,089bp to 44,721,236bp on chromosome 4. Following the sequencing data analysis we regarded this 12.1Kb deletion as a strong candidate dwarf mutation in domestic rabbits because it disrupts the coding sequence of *HMGA2*. Based on the hypothesis that the identified deletion was the single autosomal recessive dwarf mutation (d) mentioned above, genotyping on this deletion was done. We proved that our deletion genotyping was perfectly consistent with this deletion being the casual mutation according to expectations. Peanuts were homozygous for the deletion, dwarfs were heterozygous for the deletion and wild-type rabbits were homozygous without the deletion.

Except for the dwarf gene (d), an alternative gene which was designed as *diminutive* gene (b) by Greene in 1940 was presented to be a co-effective gene or a modifier which might influence the phenotypic expression of the dwarf mutation in domestic rabbits (Greene 1940). It’s effect was revealed when Polish dwarf rabbits were mated with normal unrelated animals of any line except for lines carrying the cretinoid abnormality (Greene 1940). In this study, when the rabbits are homozygous for the dwarf gene (d), they are Peanuts regardless of the diminutive genotype (b). When the rabbits are homozygous for the wild type dwarf gene (D), they are wild-type normal rabbits regardless of diminutive genotype. However, when the rabbits are heterozygous for the dwarf gene, animals of genotype Ddbb are dwarfs while DdBb animals are normal phenotype. Further analyses may be conducted to test whether the diminutive gene plays a role only in the

mentioned cross or show a similar effect on other genetic backgrounds.

The 12.1kb deletion starts 6.1 kb of upstream of exon 1 and ends at 1.1kb of downstream of exon 3 of *HMGA2* gene. This, the promoter region and the first three exons of *HMGA2* gene are deleted by this mutation which must lead to an inactivation of *HMGA2* gene. Previous studies showed that mutations in the *HMGA2* gene are associated with in reduced body weight in adults and overall decreased body size except the head in pygmy mice and in White Leghorn chicken (Ruyter-Spira *et al.* 1998). The mouse study demonstrated that inactivation of the *HMGA2* gene was responsible for the dwarf phenotype and we therefore conclude that the 12.1 deletion affecting the *HMGA2* gene is a very likely causative mutation for the dwarf and peanut phenotypes in domestic rabbits. Besides, the inactivation of *HMGA2* in mice resulting in a skeletal muscle dysfunction could provide a reliable explanation why peanuts were not viable for more than a few days post-partum. It is possible to speculate that peanuts might die of developmental retardation for muscle and skeleton by the inactivation of *HMGA2* gene.

In this study, we checked *HMGA2* expression levels on wild-type rabbit embryos at different developmental stages to test our hypothesis that *HMGA2* gene is responsible for dwarf phenotype in domestic rabbits. *HMGA2* gene has high expression level during early embryo developmental stages from day 9.5 to day 18 while dramatically decreasing on day 21 and in the following stages. In pygmy mice, there existed high expression level of *HMGA2* gene on early embryo developmental stages from day 10.5 to day 14.5 but barely detectable in adult tissues, suggesting that the HMGA2 protein is crucial for growth regulation during embryonic development (Zhou *et al.* 1995; Lee & Dutta 2007). Our expression profile analyses indicated that the *HMGA2* gene in rabbits has a very similar expression profile during development.

The analysis of *HMGA2* expression was further extended by its localization in the wild-type embryos at different developmental stages. Expression profile analysis was done on different body parts of embryos at day 21, day 24 and new born rabbit and indicating a rough order as body > head skin > skull > brain. Extremely low expression levels were observed in different parts of the head. This expression pattern was analogous with previous studies demonstrated that most tissues in pygmy mouse were 40% to 50% smaller than wild-type mouse tissues except for the brain tissue which maintained its normal size (Zhou *et al.* 1995). It might be possible to explain the appearance of extremely large head and tiny body in peanuts. However, because of the tiny size of embryos before day 15.5, we cannot dissect the embryos for special parts. The entire embryos were used as body tissue in this study. If more technologies and instruments can be used to dissect different organs or tissues of embryos at early development stage, we could compare the expression level of *HMGA2* gene in different tissues in the future.

Our results demonstrate that the absence of *HMGA2* causes growth retardation and the dwarf

phenotype in domestic rabbits. Although the precise molecular mechanism is not yet known, the function of HMGA2 protein as a transcription factor suggests that a disruption of this gene will have many downstream effects. In fact, the dwarf rabbit now provides a valuable model to further study the functional significance of the HMGA2 gene which is till poorly characterized. Based on previous studies in other species, we could test whether the *let-7* microRNA is regulating *HMGA2* expression as observed in other species. Besides, we could use RNA-seq to investigate the regulation of downstream genes of *HMGA2*. As *HMGA2* is a transcriptional factor, changes on *HMGA2* will influence other genes and consequently cause diseases or aberrant phenotypes. ChiP-seq and RNA-seq can be utilized to perform molecular genetic studies to reveal the molecular consequences of the disruption of *HMGA2* expression.

REFERENCES

Airey J.A., Baring M.D., Beck C.F., Chelliah Y., Deerinck T.J., Ellisman M.H., Houenou L.J., McKemy D.D., Sutko J.L. & Talvenheimo J. (1993) Failure to make normal α ryanodine receptor is an early event associated with the crooked neck dwarf (cn) mutation in chicken. *Developmental dynamics* **197**, 169-88.

Alyaqoub F., Pyatt R.E., Bailes A., Brock A., Deeg C., McKinney A., Astbury C., Reshmi S., Shane K.P., Thrush D.L., Sommer A. & Gastier-Foster J.M. (2012) 12q14 microdeletion associated with *HMGA2* gene disruption and growth restriction. *Am J Med Genet A* **158A**, 2925-30.

Benson K.F. & Chada K. (1994) Mini-mouse: phenotypic characterization of a transgenic insertional mutant allelic to pygmy. *Genetical research* **64**, 27-33.

Brants J.R., Ayoubi T.A., Chada K., Marchal K., Van de Ven W.J. & Petit M.M. (2004) Differential regulation of the insulin-like growth factor II mRNA-binding protein genes by architectural transcription factor *HMGA2*. *FEBS letters* **569**, 277-83.

Castle W. & Sawin P. (1941) Genetic linkage in the rabbit. *Proceedings of the National Academy of Sciences of the United States of America* **27**, 519.

Greene H.S. (1940) A DWARF MUTATION IN THE RABBIT : THE CONSTITUTIONAL INFLUENCE ON HOMOZYGOUS AND HETEROZYGOUS INDIVIDUALS. *J Exp Med* **71**, 839-56.

Greene H.S., Hu C.K. & Brown W.H. (1934) A LETHAL DWARF MUTATION IN THE RABBIT WITH STIGMATA OF ENDOCRINE ABNORMALITY. *Science* **79**, 487-8.

Hillier L.W., Marth G.T., Quinlan A.R., Dooling D., Fewell G., Barnett D., Fox P., Glasscock J.I., Hickenbotham M. & Huang W. (2008) Whole-genome sequencing and variant discovery in C. elegans. *Nature methods* **5**, 183-8.

Huang X., Feng Q., Qian Q., Zhao Q., Wang L., Wang A., Guan J., Fan D., Weng Q. & Huang T. (2009) High-throughput genotyping by whole-genome resequencing. *Genome research* **19**, 1068-76.

Kofler R., Orozco-terWengel P., De Maio N., Pandey R.V., Nolte V., Futschik A., Kosiol C. & Schlötterer C. (2011) PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PloS one* **6**, e15925.

Lee Y.S. & Dutta A. (2007) The tumor suppressor microRNA let-7 represses the *HMGA2* oncogene. *Genes & development* **21**, 1025-30.

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

Li Z., Gilbert J.A., Zhang Y., Zhang M., Qiu Q., Ramanujan K., Shavlakadze T., Eash J.K., Scaramozza A. & Goddeeris M.M. (2012) An *HMGA2*-IGF2BP2 axis regulates myoblast proliferation and myogenesis. *Developmental cell* **23**, 1176-88.

Mardis E.R. (2008) The impact of next-generation sequencing technology on genetics. *Trends in genetics* **24**, 133-41.

Mayr C., Hemann M.T. & Bartel D.P. (2007) Disrupting the pairing between let-7 and *HMGA2* enhances oncogenic transformation. *Science* **315**, 1576-9.

Meuwissen T. & Goddard M. (2010) Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics* **185**, 623-31.

Miyazawa J., Mitoro A., Kawashiri S., Chada K.K. & Imai K. (2004) Expression of mesenchyme-specific gene *HMGA2* in squamous cell carcinomas of the oral cavity. *Cancer research* **64**, 2024-9.

Morishita A., Zaidi M.R., Mitoro A., Sankarasharma D., Szabolcs M., Okada Y., D'Armiento J. & Chada K. (2013) *HMGA2* is a Driver of Tumor Metastasis. *Cancer research* **73**, 4289-99.

Raskin L., Fullen D.R., Giordano T.J., Thomas D.G., Frohm M.L., Cha K.B., Ahn J., Mukherjee B., Johnson T.M. & Gruber S.B. (2013) Transcriptome profiling identifies *HMGA2* as a biomarker of melanoma progression and prognosis. *Journal of Investigative Dermatology* **133**, 2585-92.

Reis-Filho J.S. (2009) Next-generation sequencing. *Breast Cancer Res* **11**, S12.

Ruyter-Spira C., de Groof A., van der Poel J.J., Herbergs J., Masabanda J., Fries R. & Groenen M. (1998) The HMGI-C gene is a likely candidate for the autosomal dwarf locus in the chicken. *Journal of Heredity* **89**, 295-300.

Schuster S.C. (2007) Next-generation sequencing transforms today's biology. *Nature* **200**.

Sgarra R., Rustighi A., Tessari M.A., Di Bernardo J., Altamura S., Fusco A., Manfioletti G. & Giancotti V. (2004) Nuclear phosphoproteins HMGA and their relationship with chromatin structure and cancer. *FEBS letters* **574**, 1-8.

Tay Y., Peter S., Rigoutsos I., Barahona P., Ahmed S. & Dröge P. (2009) Insights into the regulation of a common variant of *HMGA2* associated with human height during embryonic development. *Stem Cell Reviews and Reports* **5**, 328-33.

Thomas C.E., Ehrhardt A. & Kay M.A. (2003) Progress and problems with the use of viral vectors for gene therapy. *Nature Reviews Genetics* **4**, 346-58.

Thorvaldsdóttir H., Robinson J.T. & Mesirov J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* **14**, 178-92.

Thuault S., Valcourt U., Petersen M., Manfioletti G., Heldin C.-H. & Moustakas A. (2006) Transforming growth factor-β employs *HMGA2* to elicit epithelial–mesenchymal transition. *The Journal of cell biology* **174**, 175-83.

Zhou X., Benson K.F., Ashar H.R. & Chada K. (1995) Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. *Nature* **376**, 771-4.

Zhou X., Benson K.F., Przybysz K., Liu J., Cherath L., Hou Y. & Chada K. (1996) Genomic structure and expression of the murine Hmgi-c gene. *Nucleic acids research* **24**, 4071-7.

ACKNOWLEDGEMENTS

First and foremost my gratitude to all my supervisors Leif Andersson, Congying Chen and Carl-Johan Rubin, who guided and supported me during my thesis project. Thanks for their patience when I have troubles and special thanks for their sharing knowledge and interesting ideas with me. Thanks to Miguel Carneiro for valuable sample contribution and help with preparing figures used in my thesis project. Thanks for his help in teaching me the steps used for the bioinformatics analysis.

My work would not be imaginable without all my colleges and team members in IMBIM laboratory. Thanks for their help with practical advices and new ideas.

Last but not least, I thank to God and my family along with me and believe on me all the time.

APPENDIX

Table 3: Detailed information about experiment animals

Sample ID	Phenotype	Breed	Sex	Genotype	Amplicons results
ND A1	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A2	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A3	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A5	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A6	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A7	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A8	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A13	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A14	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A18	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A19	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A21	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A24	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A25	Dwarf	Netherland Dwarf	M	Del/WT	293bp+244bp
ND A26	Dwarf	Netherland Dwarf	M	Del/WT	293bp+244bp
ND A36	Dwarf	Netherland Dwarf	F	Del/WT	293bp+244bp
ND A38	Dwarf	Netherland Dwarf	M	Del/WT	293bp+244bp
ND A39	Dwarf	Netherland Dwarf	M	Del/WT	293bp+244bp
ND A41	Dwarf	Netherland Dwarf	M	Del/WT	293bp+244bp
ND A50	Dwarf	Netherland Dwarf	M	Del/WT	293bp+244bp
PNT1	Peanut	N/A	M	Del/Del	530bp
PNT2	Peanut	N/A	M	Del/Del	530bp
PNT3	Peanut	N/A	F	Del/Del	530bp
PNT5	Peanut	N/A	F	Del/Del	530bp
PNT6	Peanut	N/A	M	Del/Del	530bp
PNT7	Peanut	N/A	F	Del/Del	530bp
PNT8	Peanut	N/A	F	Del/Del	530bp
PNT9	Peanut	N/A	M	Del/Del	530bp
PNT10	Peanut	N/A	F	Del/Del	530bp
PNT11	Peanut	N/A	F	Del/Del	530bp
PNT12	Peanut	N/A	M	Del/Del	530bp
PNT13	Peanut	N/A	M	Del/Del	530bp

PNT14	Peanut	N/A	F	Del/Del	530bp
PNT15	Peanut	N/A	F	Del/Del	530bp
PST-7	Wild-type	Porto Santo	N/A	WT/WT	244bp
PST-10	Wild-type	Porto Santo	N/A	WT/WT	244bp
PST-16	Wild-type	Porto Santo	N/A	WT/WT	244bp
AC-Mu17	Wild-type	Champagne	N/A	WT/WT	244bp
AC-Mu12	Wild-type	Champagne	N/A	WT/WT	244bp
TolJun94-74	Wild-type	Toledo	N/A	WT/WT	244bp
TolJun94-103	Wild-type	Toledo	N/A	WT/WT	244bp
TolJun95-8	Wild-type	Toledo	N/A	WT/WT	244bp
Cau-5	Wild-type	Caumont	N/A	WT/WT	244bp
Cau-6	Wild-type	Caumont	N/A	WT/WT	244bp
Cau-7	Wild-type	Caumont	N/A	WT/WT	244bp
Ville-7	Wild-type	Villemolaque	N/A	WT/WT	244bp
Ville-8	Wild-type	Villemolaque	N/A	WT/WT	244bp
Ville-9	Wild-type	Villemolaque	N/A	WT/WT	244bp
Ville-10	Wild-type	Villemolaque	N/A	WT/WT	244bp
Calzada-7	Wild-type	Calzada	N/A	WT/WT	244bp
Calzada-8	Wild-type	Calzada	N/A	WT/WT	244bp
Calzada-15	Wild-type	Calzada	N/A	WT/WT	244bp
Calzada-16	Wild-type	Calzada	N/A	WT/WT	244bp

Table 4: Primer and probe information used in the experiment

Name	Sequence
Primer 1	AACCACTGGTCCTGTTCCTC
Primer 2	TAAACTCCAAGAAGGCATTCAG
Primer 3	TGGCTTAGGTAGTTGGGTCA
Primer 4	TTATGTGGCTTTCTCTGTAAC
Primer 5	CCAAGATATTTCTTTGCCATCTC
R-Primer	AAAGCAGAAGCCACTGGAGA
F-Primer	CTGTGAGGACGTCTCTTCCG
Probe	ACCTAGGAAATGGCCACAAC