Characterization of genes involved in recent adaptation

I n a u g u r a l - D i s s e r t a t i o n

zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Tobias Johannes Adolf Josef Heinen

aus Köln

Köln, 2008

Berichterstatter	**Prof. Dr. Diethard Tautz**
	Prof. Dr. Peter Nürnberg
Tag der mündlichen Prüfung	**17.11.2008**

"Es genügt nicht nur, im Allgemeinen von den Naturdingen ein Wissen zu besitzen, sondern wir müssen jedes Naturding danach untersuchen, wie es sich in seiner eigentümlichen Natur verhält. Die Naturwissenschaft hat nicht zum Ziel, das Tatsächliche zu berichten und einfach hinzunehmen, sondern vielmehr die Ursachen im Naturgeschehen zu ergründen"

Albertus Magnus, 13. Jahrhundert

Zusammenfassung

Hintergrund: Die genetische Ursache adaptiver Evolution liegt in zufälligen Mutationen begründet. Besonders Mutationen in cis-regulatorischen Bereichen von Genen sind im Zusammenhang mit positiver Selektion von großer Bedeutung. Wildlebende Populationen der Hausmaus (*Mus musculus*) stellen ein geeignetes Modellsystem dar, um die molekularen Mechanismen von Adaptationsprozessen zu untersuchen. Zwei Unterarten der Hausmaus, *Mus musculus musculus* und *Mus musculus domesticus*, prägen in Mitteleuropa eine sekundäre Kontaktzone aus, in der Hybridbildung und daraus resultierende Unfruchtbarkeit vorkommen. Deswegen sind Testis-Gene interessante Beispiele, anhand derer man Artbildungs- und Adaptationsprozesse in beiden Unterarten erforschen kann.

Vergleichende Studien in beiden Unterarten zu Genexpressionscharakteristika und um Adaptationsgene zu identifizieren wurden bereits durchgeführt. Dabei stellte sich heraus, dass die Mitogen-Aktivierte-Proteinkinase-Kinase 7 (Mkk7) und das Orphan-Gen LP10 zwei Gene sind, die sowohl an jüngst stattgefundenen Adaptationsprozessen beteiligt sind, als auch im Testis beider Unterarten differenziell exprimiert werden. Die vorliegende Arbeit beschreibt eine molekulargenetische Analyse beider Gene, die als ein langfristiger Beitrag zur experimentellen Validierung natürlicher Selektion angesehen werden kann.

Vorgehensweise: RACE, Northern-Blot und qRT-PCR Methoden wurden angewendet, um die komplizierten Expressionscharakteristika verschiedener Mkk7-Isoformen in Wildmäusen zu entschlüsseln. Die Testis-Expressionsmuster von Mkk7 wurden in *M. m. musculus* und *M. m. domesticus* mit Hilfe von in situ Hybridisierung untersucht. Die Sequenzierung genomischer DNS in beiden Unterarten führte zu der Entdeckung von Mutationen, die möglicherweise eine Ursache für die unterschiedliche Expression darstellt. Es wurde ein Luciferase-Expressions-System entwickelt, um die Aktivität solcher natürlichen Varianten von Promotoren messen zu können. Anhand von

Western-Blot-Experimenten wurden Konsequenzen der unterschiedlichen Testis-Expression von Mkk7 auf Proteinebene nachvollzogen.

Die Genexpression von LP10 wurde mit Northern-Blot-Experimenten in verschieden Wildmausarten verglichen, um den evolutiven Ursprung des Genes innerhalb des Mäusestammbaums beurteilen zu können. Eine konditionale LP10 Knockout-Linie zur funktionalen Charakterisierung von LP10 in Testis wurde etabliert. Sowohl das männliche Fortpflanzungssystem, als auch Spermien wurden in LP10 Knockout-Mäusen phänotypisiert. Die Genexpression in den Knockout-Mäusen wurde mit der in Wildtyp-Mäusen durch Microarrayanalysen verglichen.

Ergebnisse: Mkk7α1 ist eine Transkript-Isoform, die spezifisch für späte Spermatiden-Stadien in *M. m. domesticus* ist. Sie kommt nicht in *M. m. musculus* vor und wird von einem alternativen Promoter exprimiert. Es konnte der Zusammenhang zwischen einer Mutation im proximalen Abschnitt dieses Promoters und einem Expressionsunterschied beschrieben werden, der in Reportergen-Experimenten in Zellkultur gemessen wurde. Überdies wurde eine potentielle Insulator-Sequenz gefunden, die die Genexpression in anderen Zelltypen außerhalb von Spermatiden reprimiert. Aufgrund der Proteinanalysen konnte das Vorhandensein einer bislang Unbekannten Funktion von Mkk7 in Testis angenommen werden. Die Translation eines neuartigen Peptids wurde postuliert.

Es stellte sich heraus, dass LP10 im Stammbaum der Mäuse de novo aus nichtkodierender Sequenz hervorgegangen ist. Eine orthologe Region ist zwar auch in anderen Säugetieren vorhanden, wird aber nicht transkribiert. Dieses Ergebnis zeigt, dass sogar die de novo Formation von Multi-Exon-Genen möglich ist, ohne dass Genduplikation involviert ist. Der Testis-spezifische Knockout von LP10 resultiert nicht in Unfruchtbarkeit, aber vorläufige Ergebnisse deuten daraufhin, dass LP10 einen Effekt auf Testis-Gewicht und Spermienmotilität haben könnte. Die Microarrayanalysen lassen vermuten, dass LP10 die Expression einiger anderer Gene beeinflusst.

Abstract

Background: Random mutations in natural populations are the genetic basis of adaptive evolution. In particular, it is thought that mutations in cis-regulatory regions of genes are important targets for positive selection. Wild populations of house mouse subspecies (*Mus musculus*) are a convenient model to study molecular mechanisms of adaptation. The two subspecies *Mus musculus musculus* and *Mus musculus domesticus* form a zone of secondary contact in middle Europe. In this contact zone hybrid formation as well as hybrid infertility have been documented. Therefore testis genes are an interesting subject for investigating ongoing speciation and adaptation processes in both subspecies.

Surveys to compare expression signatures in both subspecies and to find adaptive trait genes were performed previously. Mitogen activated protein kinase kinase 7 (Mkk7) and the orphan gene LP10, were identified as adaptive trait genes differentially expressed in testis of *Mus musculus musculus* and *Mus musculus domesticus*. The work presented here describes a molecular genetic analysis of both genes as part of a long term approach to provide experimental evidence for natural selection.

Approach: RACE, Northern blot and qRT-PCR experiments were performed to resolve the details of the complex expression pattern of different isoforms of Mkk7 in wild mice. Testis expression of Mkk7 in *M. m. musculus* and *M. m. domesticus* was examined by situ hybridization. Genomic sequencing of both subspecies identified mutations possibly contributing to cis regulatory differences. A luciferase expression system was developed for testing expression strength of Mkk7 promoter variants. Western blot experiments were performed to follow consequences of Mkk7 expression changes at the protein level.

LP10 expression in different mouse species was assessed in Northern blot experiments to track the origin of the gene in the mouse lineage. A LP10 conditional knock-out mouse line was established to study LP10 function in testis. The male reproductive

organs and spermatozoa of LP10 knock-out mice were phenotyped. Testis expression of knock-out mice and wild-type mice was compared by microarray analysis.

Results: The Mkk7α1 transcript is an isoform specifically expressed in late spermatids in *M. m. domesticus*. It does not occur in *M. m. musculus* and is driven by an alternative promoter. A single proximal mutation in this promoter was identified as a possible cause of expression differences in cell culture reporter gene experiments. A potential insulator sequence was found that represses expression in non-testis cells. Evidence for a so far unknown function of Mkk7 in testis is deduced from protein analysis. The translation of a novel peptide is postulated.

It appears that LP10 has evolved de novo from a non-coding sequence in the mouse lineage. An orthologous region is present in other mammals, but is not transcribed. This result show that even a locus with multiple exons can form de novo, without any gene duplications involved. The knockout of LP10 in testis does not lead to infertility, but the preliminary results suggest an effect on testis weight and sperm motility. The microarray analysis suggests that a small number of other genes are affected in LP10 knockout mice.

Content

1 Introduction

How random mutations in the genomes of organisms lead to adaptation and speciation is of great interest to evolutionary biologists. To compliment genome wide surveys exploring general mechanisms of evolution, understanding the contribution of changes in individual genes to adaptation and speciation is also of great importance. In a growing number of cases it has been possible to decipher the role of individual genes in adaptations. One of the classic examples is the radiation of Darwin's finches. Differences in the development of beak shapes were found to be related to differences in expression levels of the calmodulin gene in beak primordia of Darwin's finches (Abzhanov et al., 2006). Warfarin resistance in rat evolved through selection of mutations in a vitamin K reductase gene (Kohn et al., 2000, 2003; Pelz et al., 2005). In humans, the beta-globin gene which is target of selection related to malaria resistance (Hanchard et al., 2007) and the FOXP2 gene which is involved in the evolution of speech (Enard et al., 2002) are two prominent examples. However, in order to build a more comprehensive understanding of evolutionary mechanisms it will be a future challenge for biologists to dissect more examples of such genotype-phenotype relationships in wild populations in a functional and ecological context.

1.1 Evolutionary biology in the era of modern genomics

A reasonable approach to investigate the genetics of evolutionary changes is the comparison of wild populations of closely related species, subspecies or subpopulations. The era of modern genomics provides new technologies that facilitate such comparisons (Ellegren and Sheldon, 2008). Acquisition of large amounts of genomic sequence data from wild population samples brings new insights into genome evolution. It also offers the possibility to screen wild populations for so called selective sweep loci. These are genomic regions containing a beneficial mutation that has been target of natural selection leading to fixation of the allele and its linked flanking regions – a process called genetic hitchhiking which results in reduced sequence variability in the vicinity of the sweep (Smith and Haigh, 1974). Selective sweeps can be identified for example

via microsatellite analysis (Ihle et al., 2006; Kauer et al., 2003; Wiehe et al., 2007) or with statistical tests based on genomic sequence data (Fu, 1997; Kim and Stephan, 2002; Tajima, 1989). The invention of microarray technologies allowed the consideration of large amounts of gene expression data. Measuring the expression differences between closely related species can establish a basis for understanding the contribution of expression variation to evolutionary processes (Rottscheidt and Harr, 2007; Voolstra et al., 2007). In the near future, next generation sequencing will gather importance in expression profiling (Torres et al., 2008) and will open up the possibility to compare sequence and expression data, also in non-model organisms and samples of wild populations.

1.2 The house mouse as a model in evolutionary biology

Among mammals, the house mouse *Mus musculus* (LINNAEUS, 1758) is the best established model organism. It offers a lot of genomic resources, it serves as a genetic model for human traits and, in particular, it provides a large variety of fundamental and unique tools for experimental application, such as homologous recombination in embryonic stem cells. In addition to the research that has been carried out using inbred strains, studies involving wild mice have amassed basic knowledge about the ecology and population structure of this species making it also an attractive model for evolutionary studies.

The taxon *Mus* is split into different sub-taxa (Chevret et al., 2005), which have spread over the entire globe. The house mouse lineage itself consists of different subgroups. This phylogenetic diversification is believed to be the result of a radiation that took place in the Northern part of the Indian sub-continent (Boursot et al., 1993). The three best characterized subgroups within the *Mus musculus* lineage are *Mus musculus domesticus* (Western Europe and areas that have been colonized by Western European nations), *Mus musculus musculus* (Eastern Europe and Northern Asia) and *Mus musculus castaneus* (Southeast Asia). These groups can also be seen as the progenitors of the common laboratory strains, which are mostly genetic mixtures of all three

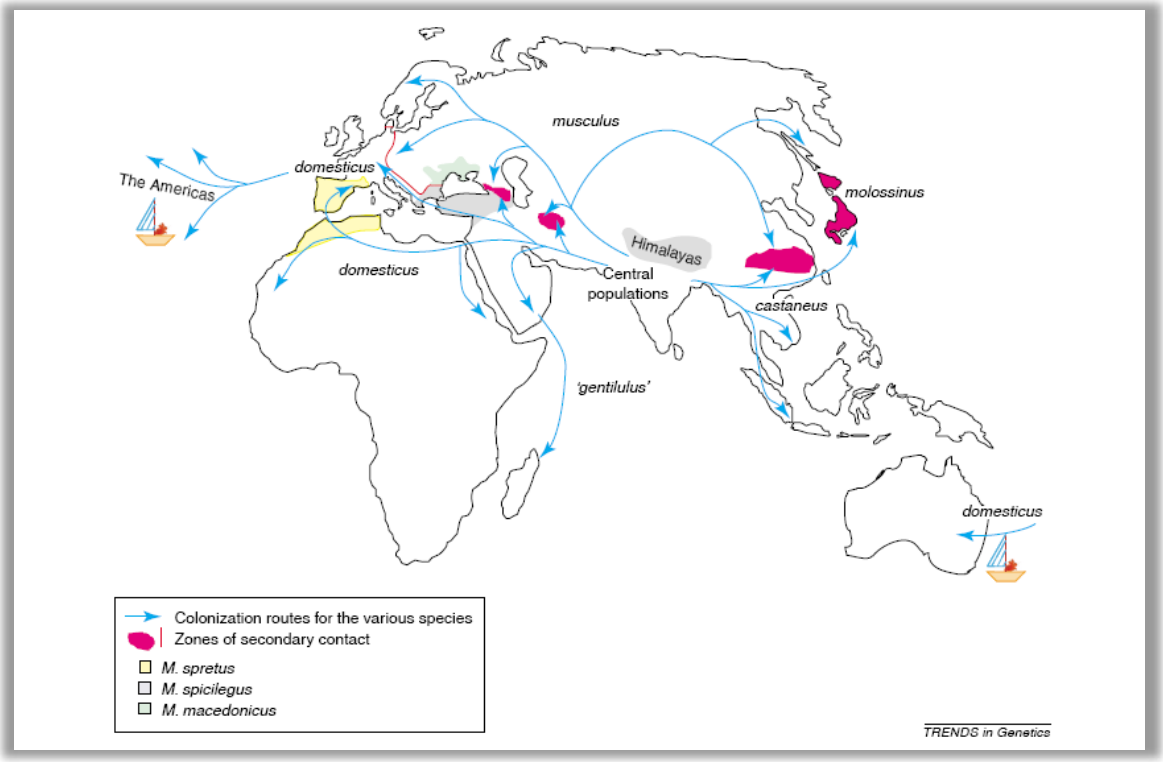


Figure 1 Geographical distribution and routes of colonization of different mouse species and subspecies. The illustration is an original picture from "Trends in Genetics" (Guenet and Bonhomme, 2003).

subspecies (Bonhomme et al., 1987). Questions regarding the origin and the ancestral populations of *Mus musculus* and questions of whether the subpopulations should be regarded as distinct species or as different subspecies are still the subjects of ongoing debate. In the following, the subgroups of *Mus musculus* will be called subspecies as commonly used in the literature.

Figure 1 illustrates the geographical distribution and colonization routes of the main subspecies. The different subspecies are separated by contact zones in which hybrid formation occurs. A well-defined hybrid zone between mouse subspecies is formed by *M. m domesticus* and *M. m. musculus* (Hunt and Selander, 1973; Sage et al., 1986; Teeter et al., 2008). Both lineages were separated roughly 0.9 million years ago (Boursot et al., 1996) and formed a zone of secondary contact across Central Europe 3000 years ago (Cucchi et al., 2005). This hybrid zone reaches from Denmark to the Black sea and crosses Germany, the Czech Republic, Austria and the Balkans. Among

hybrids between *M. m. musculus* and *M. m. domesticus* a certain degree of hybrid sterility can be observed in wild mice and inbred strains (Britton-Davidian et al., 2005; Forejt, 1996). Nevertheless, the hybrid zone is characterized by a narrow cline and gene introgression in both directions has been reported, but some regions of the genome, particularly sex chromosome markers, introgress less fast and less far suggesting the existence of incompatibility factors involved in reproductive isolation and speciation (Macholan et al., 2007a; Teeter et al., 2008; Tucker et al., 1992). A recent study supports this theory by showing that fertilization occurs significantly faster in conspecific versus heterospecific matings and that *M. m. musculus* sperm consistently out-competed *M. m. domesticus* sperm (Dean and Nachman, 2008).

1.3 Aim of the thesis

This thesis presents different examples of potentially adaptive trait genes in house mice. All projects focus on genes that are differentially expressed between wild populations of *Mus musculus*. In case that those expression differences are caused by cis-effects, the advantage of investigating them instead of protein mutations is the availability of experimental strategies to measure gene expression. The disadvantage is the poor understanding and characterization of cis regions in comparison to coding regions. Reporter constructs can help to relate genotype and phenotype, by identifying the regulatory sequences that are responsible for the expression differences (McGregor et al., 2007).

With regard to the reproductive isolation of *M. m. musculus* and *M. m. domesticus* and to other sex related processes like sexual selection and adaptation through sperm competition which presumably occurred within the two lineages, it is informative to look at gene expression levels in the testes of both subspecies. The projects described in this thesis are embedded in a series of studies to investigate the contribution of changes in gene expression to adaptation. As shown in Figure 2, one long term goal of the overall approach is to define the differences of an adaptive trait gene between both

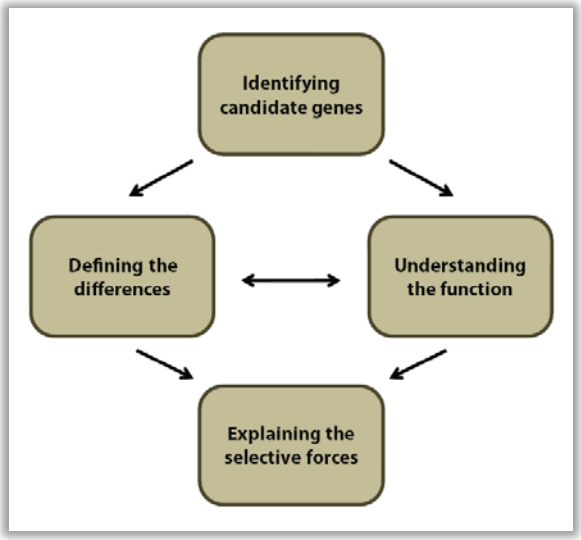


Figure 2 Approach and long term goal

subspecies and explain the functional consequences of the difference in order to understand the selective forces that led to the differences.

Two parts of the thesis will each describe a gene that was found to be differentially expressed in previous microarray analyses of testes from *M. m. musculus* and *M. m. domesticus* (MAP kinase kinase 7 and LP10). Both loci also display strong evidence of a selective sweep in one of the subspecies. They were chosen for further analysis in an unbiased way, according to their significance in expression difference and nucleotide polymorphisms, and not according to their function in testis, which remains unclear. The aim of the two projects was to perform molecular and functional characterization of both loci and to demonstrate how the technological tools that are available for mouse can be used to help answer evolutionary questions in a functional context.

Map kinase kinase 7 is a well known gene and is extensively described in the literature. A characteristic of this gene is its complex regulation and the existence of different isoforms. The work presented in this thesis focuses on analyzing its regulation in testis. RACE experiments, Northern blotting and quantitative real time PCR were used to define different isoforms and their different expression behavior in wild mice. Genomic sequencing identified mutations possibly involved in cis-regulatory differences. A reporter system was developed to test their influence on gene expression. Western blot experiments were performed to estimate the consequences of Map kinase kinase 7 expression on protein level in wild mice.

In contrast to Map kinase kinase 7, nothing is known about regulation and function of LP10. A conditional knock-out mouse line of LP10 is presented. Phenotypic analysis of

LP10 knock-out mice was done to understand its role in the male reproductive system. Microarray analysis comparing wild type and LP10 knock-out mice was addressed to discover potential influence on regulation of other genes. Northern blot experiments lead to new insights into the evolution of LP10.

Finally, the results are discussed with regard to the general approach to find experimental evidence for natural selection.

2 Mitogen activated protein kinase kinase 7

2.1 Mkk7 and MAP-kinase cascades

Mitogen activated protein kinase (MAPK) pathways are conserved eukaryotic multistep signaling cascades mediating transcriptional response upon reception of certain extracellular stimuli (English et al., 1999). They are composed of different kinases which activate their downstream targets through phosphorylation. MAPK cascades contain at least three kinases: MAP kinase kinase kinases (MKKK), which phosphorylate MAP kinase kinases (MKK), which can activate the MAPKs. There are numerous different MAPK pathways that can be interlinked in a complex way and contribute to many important cellular functions (Chang and Karin, 2001). The best characterized MAPK pathways are the extracellular signal related kinase (ERK), the p38 and the c-Jun N-terminal kinase (JNK) pathways. The murine mitogen activated protein kinase kinase 7 (MKK7) belongs to the JNK group of kinases (Holland et al., 1997; Tournier et al., 1997). It is activated upon phosphorylation of the serine and threonine residues in an S-K-A-K-T motive by MKKKs, including MEKK1 and MEKK2, after docking to a DVD-motive at the C-terminus of Mkk7 (Takekawa et al., 2005; Wang et al., 2007). As dual specificity kinases, Mkk7 and Mkk4 are both able to phosphorylate JNK at threonine and tyrosine residues in the phosphorylation motive threonine-proline-tyrosine in a synergistic manner. Mkk7 is a specific activator of JNK, but displays higher substrate specificity towards the threonine residue, while Mkk4 preferentially phosphorylates tyrosine and can also activate p38 (Fleming et al., 2000; Kishimoto et al., 2003; Tournier et al., 2001; Wang et al., 2007).

The JNK pathway responds to different stimuli. Cellular stress like UV and gamma irradiation, osmotic shock and drug treatments on the one hand and different inflammatory cytokines such as tumor necrosis factor (TNF), interleukin-1 or interleukin-3 on the other hand lead to JNK activation (Chang and Karin, 2001; Foltz et al., 1998; Moriguchi et al., 1997; Nishina et al., 2004). Downstream targets of JNK include transcription factors (Yang et al., 2003b) as well as other proteins, for example

microtubule-associated proteins (Chang et al., 2003) and members of the Bcl-2 family (Deng et al., 2003; Lei et al., 2002). The JNK pathway has several major functions in physiology and cellular processes, for example an involvement in the immune system and in apoptosis, and moreover it is also important for developmental processes (Dong et al., 2000; Nishina et al., 2004; Sabapathy et al., 1999; Wada et al., 2004; Wang et al., 2007). Thus, double mutant mice lacking the JNK1 and JNK2 isoforms and Mkk7 total knockout mice die during embryogenesis.

2.2 Mkk7 is a potential adaptive trait gene

A previous study showed that a testis specific Mkk7 expression difference between *M. m. musculus* and *M. m. domesticus* is associated with a selective sweep in *M. m. domesticus* (Harr et al., 2006). An elevated level of Mkk7 RNA in domesticus compared to *M. m. musculus* was first noticed through a microarray analysis and was confirmed by quantitative real time PCR. The signature of a selective sweep was seen after genomic sequencing of the Mkk7 locus in wild mice. The expression difference and the differences in nucleotide diversity are consistent within both subspecies. A cis-trans test via allele specific expression analysis in F1 hybrids of both subspecies demonstrated that the expression change is caused by a cis-acting sequence difference. The attempt to demonstrate the expression difference using Northern blotting revealed that an additional testis specific band is representing the higher Mkk7 transcription in *M. m. domesticus*.

These findings about Mkk7 raise the question about their cause and effect. A detailed knowledge about the molecular mechanisms leading to the expression differences and about its functional consequences might help to understand if and how the changes of Mkk7 transcription are contributing to evolutionary processes. The definition of the different Mkk7 mRNA variants in testis and the description of their expression levels are prerequisites for the formulation of further questions dealing with mechanistic aspects. Determining the transcript structure is the first step in identifying the cis-regulatory sequences underlying the subspecies specific differences, which are possibly

the target of selection and the cause of the selective sweep in the *M. m. domesticus* lineage.

Unraveling the transcriptional differences of Mkk7 between *M. m. musculus* and *M. m. domesticus* will open up the possibility to discover more about the function of Mkk7 in testis. It would be of great interest to describe the mechanisms by which Mkk7 acts in testis and particularly to know whether Mkk7 is involved in spermatogenesis. The following disquisition will approximate those questions.

2.3 Isoforms of Mkk7 transcripts in testis

2.3.1 Different Mkk7 transcripts in *M. m. musculus* and *M. m. domesticus*

To demonstrate the expression of Mkk7 in testis, total RNA from *M. m. domesticus, M. m. musculus* and from the standard laboratory strain Black6 was analyzed with Northern blotting. Figure 3A shows the result. All lanes contain two bands indicating the existence of different isoforms. One band appears at nearly 4 kb and a smaller one at roughly 1.6 kb. The larger band is stronger in *M. m. musculus*, whereas the smaller band is very strong in *M. m. domesticus* and very weak in *M. m. musculus*. The striking disparity of the smaller bands reflects the overall expression difference previously detected in the microarray and real time PCR experiments. The band pattern of the Black6 sample is consistent with the one from *M. m. domesticus*. 92% of the Black6 genome is of *M. m. domesticus* origin (Frazer et al., 2007) and the Mkk7 locus is most likely originated from *M. m. domesticus* as well. Figure 3B shows a Northern blot analysis from a prior study (Tournier et al., 1997) that revealed the widespread expression of Mkk7 among different organs in lab strain mice. In this paper, the additional testis specific band is noted but not further discussed. Taken together from the results of both Northern blots, it can be concluded that a different expression of Mkk7 isoforms restricted to testis is responsible for the expression difference between *M. m. musculus* and *M. m. domesticus*.

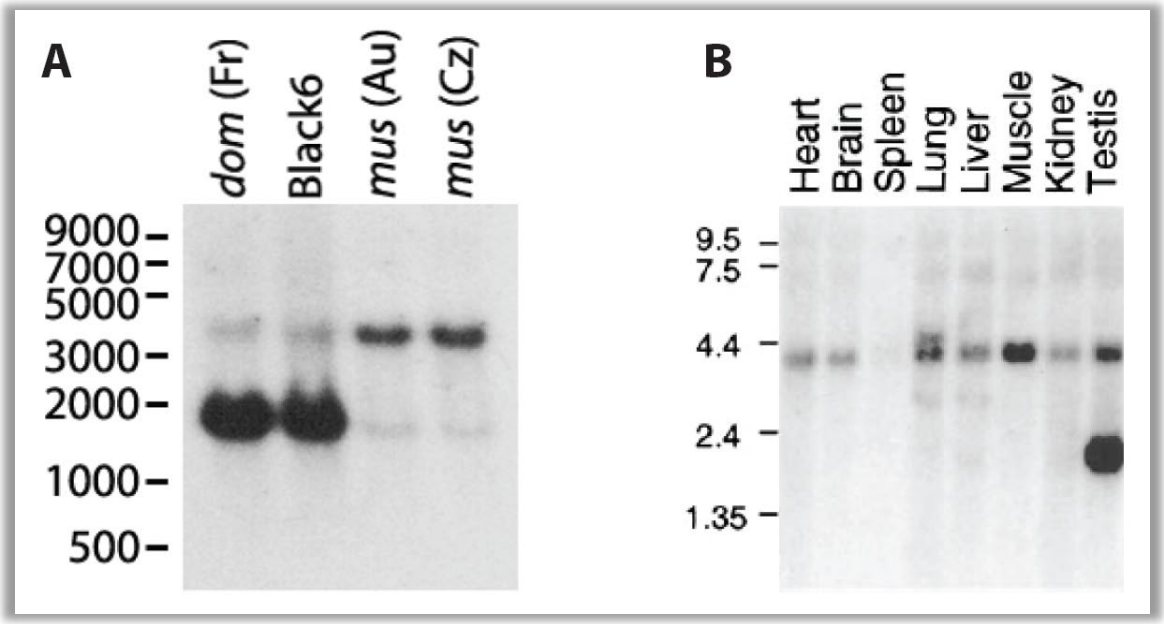


Figure 3 Mkk7 Northern blots (A) A Northern blot comparing testis total RNA of *M. m. domesticus* (from France), *M. m. musculus* (from Austria and Czech Republic) and Black6 lab strain. The blot was hybridized with a cDNA probe representing all exons that contain the kinase domain of Mkk7. Two bands are visible in each lane. The ~4kb band is stronger in *M. m. musculus*, while a ~1.6kb band is much stronger in *M. m. domesticus*. The Black6 result matches the band pattern of *M. m. domesticus*. The size standard indicates bp. (B) An Mkk7 RNA of ~4kb in size is expressed in many mouse organs. The additional 1.6kb band is testis specific. The size standard indicates kb. This picture is taken from Tournier et al. (1997).

In order to understand the cause and consequences of this phenomenon, it is necessary to define the Mkk7 isoforms that correspond to the different bands. Analysis of the annotation of the Mkk7 locus reveals that this is not a self-evident problem that can be solved intuitively, because Mkk7 is a gene locus of high complexity. Figure 4A gives an overview about the EST-genes and mRNAs that are annotated in the MGI genome browser for this locus (http://www.informatics.jax.org/). A closer look at the annotations clarifies that the length of the transcript is mainly determined by alternative splicing in the 3' region and by different transcript endings. The 5' promoter region determines the expression strength and also the expression difference between *M. m. musculus* and *M. m. domesticus* which is caused by cis acting factors (Harr et al., 2006). Unfortunately, there are different transcription start sites indicating the existence of different promoters. It is unclear which 5'- and 3'-variants are combined in testes of

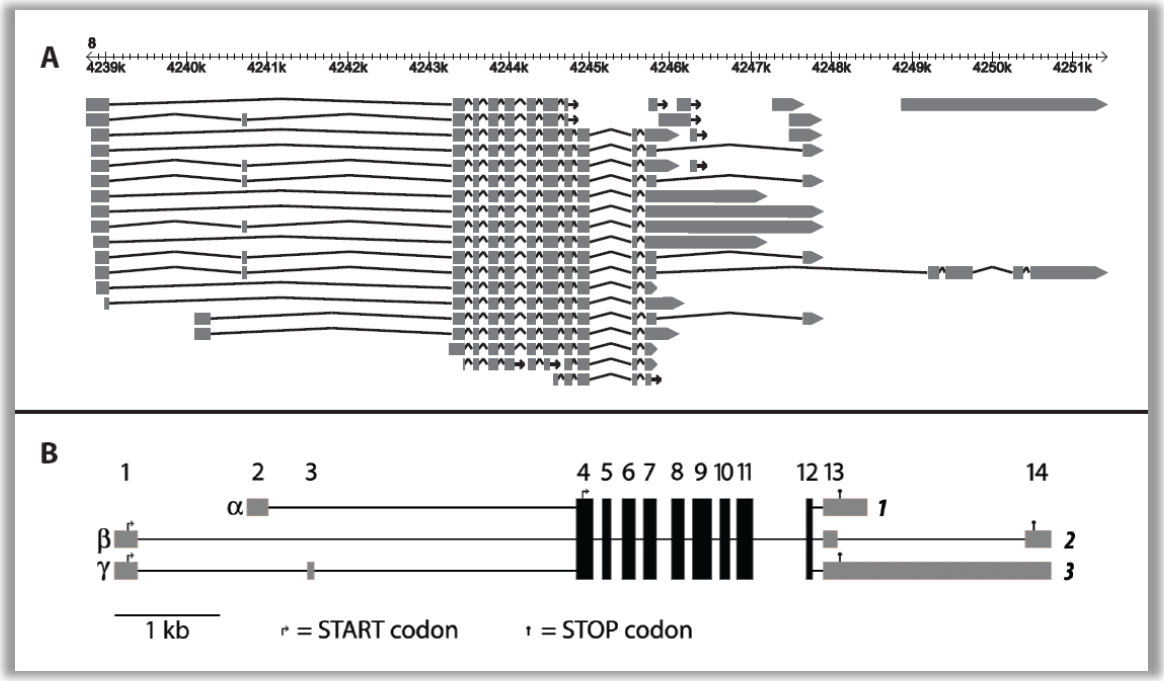


Figure 4 Exons of the murine Mkk7 locus (A) Annotation of EST-genes and cDNAs are shown in the genomic context of chromosome 8. Source: http://www.informatics.jax.org/, 2008 (B) Model of Mkk7 isoforms modified according to Tournier et al. (1999).

different wild mice and which variants account for the bands that came up in Northern blots.

Tournier and colleagues identified different isoforms of Mkk7 by screening testis cDNA clones from laboratory mice and implemented a nomenclature (Tournier et al., 1999). They found three different 5' versions (alpha, beta, gamma) and two 3' variants (1 and 2) in all possible combinations. Both 3' forms represent a short (~1.6kb) version of Mkk7. Thus, the model lacks an explanation for the long transcript; but among the annotated transcripts, there is just one 3' variant that is long enough to be the large band in the Northern blot. In the following, this variant will be called variant 3. Beta and gamma are driven by the same promoter and differ in the absence or presence of the alternatively spliced exon 3. Translation begins in both cases at a start codon in exon 1, while the presence of exon 3 does not change the reading frame. The transcription of alpha starts from a promoter located in the first intron of both beta and gamma variants

and begins with exon 2, which does not contain a start codon. An alternative start codon can be found in exon 4 leading to a shorter N-terminus. The length of the C-terminus depends on which 3' exons are used. Variants 1 and 3 generate the same C-terminus by sharing a stop codon, but variant 3 overlaps with the polyA signal of variant 1 and creates a longer 3' untranslated region (UTR) of exon 13. Variant 2 uses the same polyA signal as variant 3, but a large piece is spliced out generating an additional exon 14. In this case the splice donor lies upstream of the stop codon that is used in the variants 1 and 3. Instead, another stop codon in exon 14 is used. This leads to a longer C-terminus. The exons 4 – 12 are shared by all variants. A graphical overview of the adjusted Tournier-model is shown in Figure 4B. This nomenclature of Mkk7 isoforms serves as basis for following experiments.

In order to define different Mkk7 isoforms in wild mice RACE experiments were carried out. RACE templates were generated with total testis RNA from *M. m. musculus* and *M. m. domesticus*. In the case of complicated loci like Mkk7 there is a risk that some of the variants will be missed due to technical limitations. 5' and 3' RACE-PCRs with Mkk7 specific primers were run as gradients with different annealing temperatures to minimize this risk. A fraction of all PCRs was run on an agarose gel to inspect the PCR products. The DNA was transferred onto a membrane by Southern blotting and hybridized with an Mkk7 probe. This control experiment verifies the specificity of the PCR products and allows visualization of very week bands because of its high sensitivity. The results are shown in Figure 5. The gene specific primer used in the 5'RACE-PCR lies 784 bp downstream of exon 4. In both subspecies the 5' RACE produced a band of nearly 1 kb at any annealing temperature and an additional shorter band at low annealing temperatures (Figure 5A). The Southern blot shows that the shorter band does not belong to Mkk7, but the larger band is derived from Mkk7 transcripts. The blot also shows that Mkk7 fragments with different sizes which are too weak to be seen in the gel are present at higher annealing temperatures. Only the PCR reactions that were run with annealing temperatures of 69.4°C and 70.7°C (lanes 9 and 10) were pooled separately for *M. m. domesticus* and *M. m. musculus* to exclude a

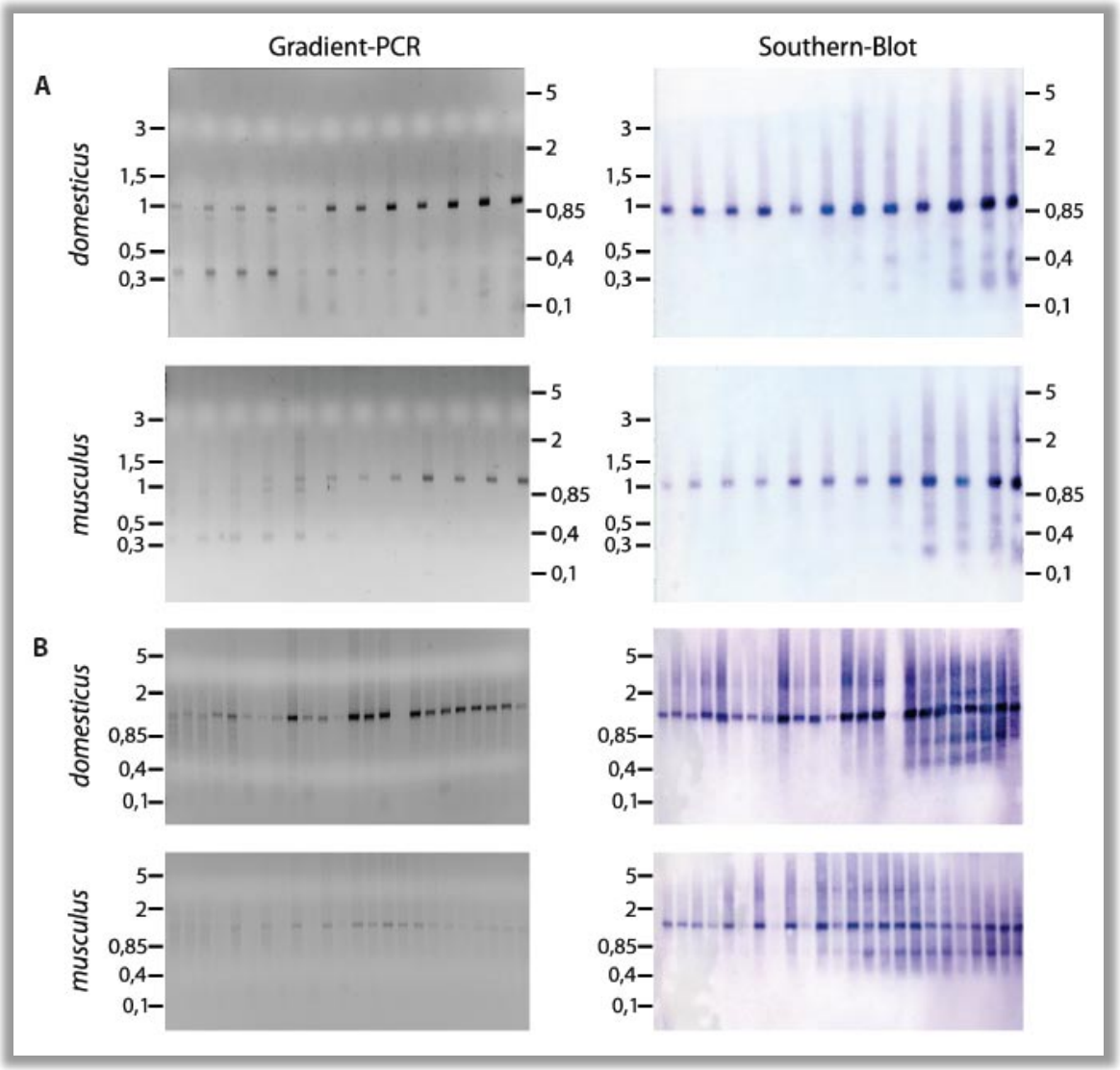


Figure 5 RACE-PCRs for *M. m. domesticus* and *M. m. musculus*. Gradient PCRs (left site) were run at following annealing temperatures (from left): 55°C, 55.4°C, 56.5°C, 57.8°C, 59.8°C, 62.2°C, 65.1°C, 68.3°C, 69.4°C, 70.7°C, 71.7°C, 72°C. PCRs were done in 10 µl reaction volumes. 1 µl per reaction was loaded on an agarose gel. The size standards indicate kb. The gradients were blotted on a membrane and Mkk7 specific bands were detected with a DIG labeled probe (right site). (A) 5' RACE was performed with 2 minutes elongation time. It shows one specific band of 1 kb and additional weak signals at high annealing temperatures. (B) 3' RACE was performed with 2 minutes elongation time (every second lane starting left-aligned) and with 4 minutes elongation time (staggered between the lanes of the first gradient starting in second right behind the first lane). It shows a specific band at roughly 1 kb and additional weak bands at high annealing temperatures. A clear band that corresponds to the long transcript was not obtained although a prolonged elongation time was tried.

disturbing amount of DNA from the unspecific band. Both pools were each cloned into a PCR cloning vector and transformed into *E. coli*. 48 clones of each subspecies were picked and sequenced for further analysis.

Figure 5B displays the results of the same experimental setup for Mkk7 3' RACE. The RACE-PCR produced one visible band at roughly 1 kb in all temperature steps of the gradient. The blot indicates that this band is indeed an Mkk7 fragment. The gene specific primer binds 862 bp upstream of exon 13 implying that the PCR band may represent one of the 3' variants 1 or 2; but it is too small to be a product of the long variant 3. Also, a PCR with prolonged elongation time does not enhance the generation of a larger RACE fragment. On the other hand, the blot reveals the existence of a number of weak intensity bands that are larger or smaller than 1 kb. Based on these results, all PCR reactions were pooled separately for *M. m. musculus* and *M. m. domesticus*. 60 RACE clones of each subspecies were obtained for sequencing as described above. The sequence data of the RACE clones can be found in the digital supplement. Table 1 summarizes the outcome of the RACE experiment.

The result corroborates the complex regulation of the Mkk7 locus. Start sites can be found within six different exons apart from the common ones. A multitude of transcript end points resulting in shorter RNAs were also detected. A broad overview of different isoforms is obtained, as a result of the cloning of PCR pools followed by the analysis of a large number of clones instead of cloning or direct sequencing of PCR bands. This method diminishes the risk of missing a certain variant, but it also increases the occurrence of non-specific PCR products, short fragments and unusual sequences such as potential trans-splicing forms. Those clones are listed as "artifacts" in Table 1. The most important outcome of the experiment is the description of common Mkk7 variants as they are shown in Figure 4B. The data indicates that at least some of the bands that were obtained in the RACE-PCRs are mixtures of different variants with same length. The alpha variant is most common among all 5' variants in *M. m. domesticus* while it is absent among the *M. m. musculus* clones. On the other hand, the beta variant is most abundant in *M. m. musculus*, but less so in *M. m. domesticus*. The gamma variant is not

5' RACE

Variants	dom	mus
alpha	7	0
beta	2	5
gamma	0	0
Other transcript start positions in:		
Exon 5	0	1
Exon 6	1	0
Exon 7	0	1
Exon 8	0	1
Exon 9	4	2
Exon 10	2	1
artifacts	32	37
Total	48	48

3' RACE

Variants	dom	mus
1	26	0
2	2	3
3	0	0
Other transcript ends in:		
Exon 9	1	1
Exon 10	0	3
Exon 12	1	0
Exon 13	3	0
artifacts	27	53
Total	60	60

Table 1 Analysis of Mkk7 5' RACE and 3' RACE clones from *M .m. domesticus* (dom) and *M. m. musculus* (mus) testis. The number of different transcript start variants (5' RACE) and transcript end variants (3' RACE) were counted. Non-specific clones or clones that cannot be identified are categorized as artifacts.

detected at all. Almost half of the 3' clones from *M. m. domesticus* correspond to variant 1, while variant 2 is rather rare. The second variant is also rare in musculus, but it is the only version that was detected in this subspecies. As expected, variant 3 could not be identified in either of the subspecies, since the PCR bands were not large enough to explain the long transcript. In order to prove the existence of variant 3 another RACE-PCR was performed.

Obviously, the 3'-RACE PCR reaction preferentially amplifies shorter transcripts in a competitive manner. Moreover the subsequent ligation of PCR pools into cloning vectors enriches short artifactual fragments. To circumvent this, a different gene specific primer placed in exon 13 was chosen. It binds downstream of the transcript end of variant 1 and upstream of exon 14 from variant 2. Thus, the amplification of both

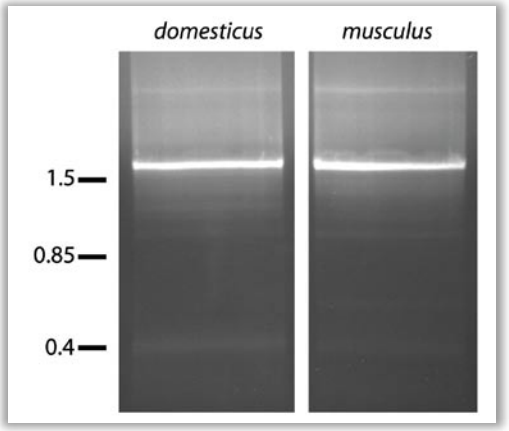


Figure 6 3' RACE gradient PCR pools from domesticus and musculus. The size standard indicates kb.

isoforms is excluded. The assumed length of the PCR product of variant 3 is 1676 bp. A gradient PCR for both subspecies was pooled and run on an agarose gel (Figure 6). The bands at roughly 1.6 kb were extracted from the gel and sequenced directly. The sequence raw data is attached in the digital supplement. It revealed that the PCR band indeed reflects variant 3. Several alternative hypotheses are raised by these RACE experiments. The length of Mkk7 transcripts is determined by alternative splicing in the 3' region and by the usage of different stop or polyadenylation sites. The short testis specific band in Mkk7 Northern blots consists of a mixture of different equally long variants. Variant 2 is an infrequent short variant in both subspecies. Mkk7α1 seems to be a *M. m. domesticus* specific short variant. It is the dominant form in *M. m. domesticus* testis and might be the reason for the expression difference. The larger band in the Northern blot can be explained with the Mkk7β3 variant which is present in *M. m. musculus* and *M. m. domesticus*. Mkk7γ seems to be rare in testis. Furthermore, a BLAST search querying the gamma associated exon 3 against the NCBI mouse RNA and EST databases does not result in any testis hits. This fact endorses the conclusion that the gamma variant does not account for a significant proportion of the total Mkk7 RNA in testis. Therefore, Mkk7γ is not considered further with regard to the expression differences between *M. m. musculus* and *M. m. domesticus*.

2.3.2 Quantification of different Mkk7 transcript variants in wild mice

RACE-PCR is a powerful tool to obtain an overview of different isoforms, but it is not useful for quantification of isoforms. The PCR reaction favors the amplification of certain fragments that are shorter than others and the subsequent ligation reaction also biases the cloning. Therefore in order to verify the hypotheses that Mkk7α1 is *M. m.*

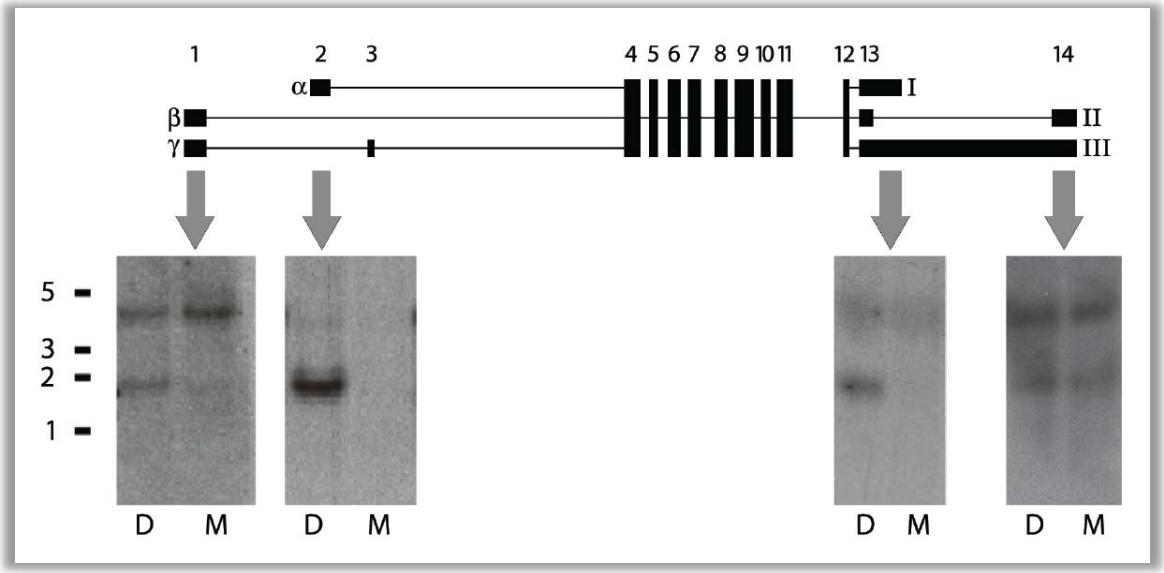


Figure 7 Exon specific MkK7 Northern blots. 10µg of testis RNA from *M. m. domesticus* (D) and *M. m. musculus* (M) is hybridized with different probes against certain parts of Mkk7. Probe positions of the respective blots are indicated with grey arrows. The size standard on the left displays kb.

domesticus specific and causes the expression difference, comparative Northern blots with probes binding to certain regions of the Mkk7 transcripts were performed (Figure 7). Hybridization with an exon 2 probe confirms that the alpha variant is highly expressed in *M. m. domesticus* while absent in *M. m. musculus*. The promoter of the alpha variant predominantly drives the expression of a short transcript. Only a very small signal of a large band can be guessed. The promoter that initiates transcription of Mkk7 beta drives the expression of long and short transcripts in both subspecies. The large band is slightly weaker in *M. m. domesticus* as can be observed in Figure 3A. In addition, the short band is stronger in *M. m. domesticus*, but the proportion of the small band in comparison to the large is far less than in Figure 3A or than the effect of the alpha variant. Another probe is designed to bind to the 3' variant 1, but not to variant 2 by placing it in exon 13 directly downstream of the splice position. It detects variant 3 equally strongly in both subspecies while variant 1 exclusively appears in *M. m. domesticus*. A probe against exon 14 hybridizes the variants 2 and 3 equally strong. From these Northern blots it can be concluded that Mkk7α1 is indeed associated with the *M. m. domesticus* lineage. The Mkk7 alpha promoter is responsible for the

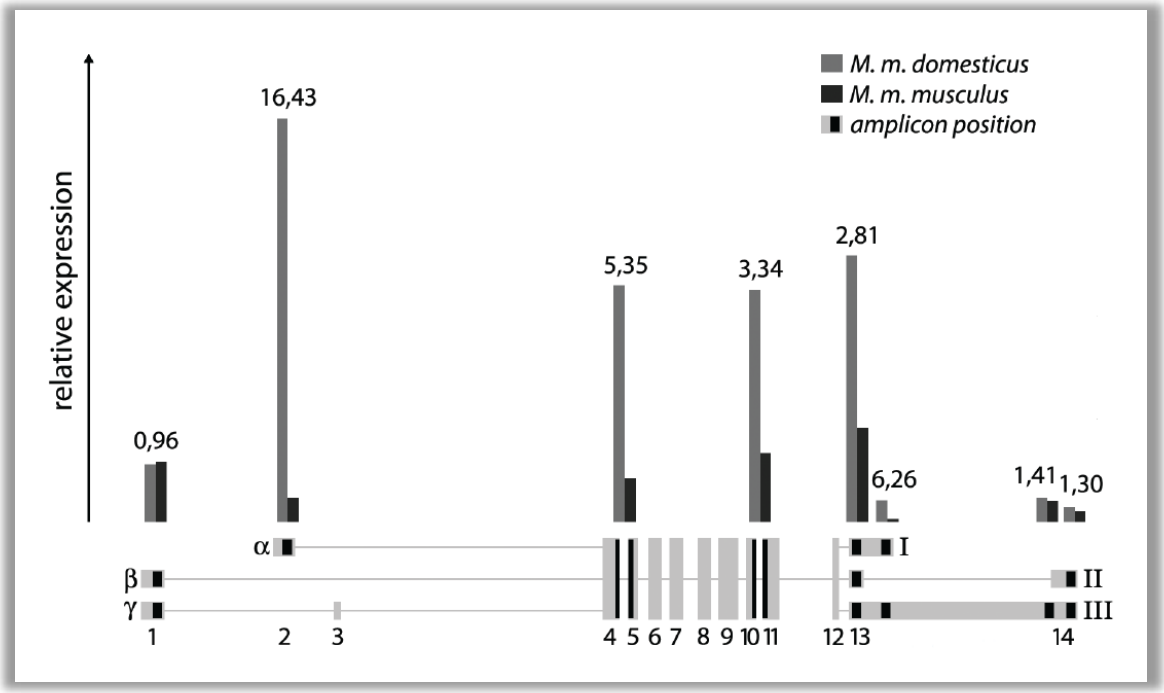


Figure 8 SYBR-green qRT-PCR of Mkk7 testis cDNA from *M. m. musculus* and *M. m. domesticus*. Relative expression of different amplicons (about 100 bp in size) is shown. Two amplicons span neighboring exons (4-5; 10-11). Positions of the different amplicons are indicated by black areas in the exon map. The respective expression values are displayed as bars above. The numbers on top of the bars represent the ratio of the domesticus to the musculus value.

difference in total Mkk7 RNA levels between *M. m. domesticus* and *M. m. musculus*. Transcripts expressed from the alpha promoter usually become short variant 1. The length of the variant is caused by the use of an early polyadenylation signal in exon 13, which is not used in *M. m. musculus*. Northern blots with probes against total Mkk7 (Figure 3A) and exon 1 (Figure 7) show a large band which is slightly weaker in *M. m. domesticus*. This can be explained by a slightly reduced activity of the Mkk7 beta promoter in *M. m. domesticus*. It can also be explained by the fact that a proportion of transcripts that exist as variant 3 in *M. m. musculus* use the early polyadenylation signal in exon 13 and are converted into variant 1 in *M. m. domesticus*.

A more precise quantification of the different isoforms can be achieved by quantitative real-time-PCR (qRT-PCR). The amplification of different Mkk7 exons from *M. m. domesticus* and *M. m. musculus* cDNA was detected with SYBR-green. The results are

shown in Figure 8. They reinforce the conclusions from RACE and Northern experiments. Exon 1 is slightly higher expressed in *M. m. musculus*. A large expression difference in exon 2 can be measured. The alpha related expression difference is maintained in the following exons, but the ratio becomes lower towards the 3' end. This can be explained with numerous transcription start sites in later exons as described in Figure 4A and Table 1. Those additional transcripts are obviously diluting the expression difference. The region between the first splice site in exon 13 and the end of variant 1 displays an exception. The amplicon at this position shows a higher ratio compared to the three amplicons in front of it. This can be explained by variant 1, which does not exist in *M. m. musculus*. Nevertheless, the qRT-PCR of exon 2 raises the question of whether the alpha promoter is really absent in *M. m. musculus*. To answer that question, a gradient PCR on 3'RACE template from *M. m. domesticus* and *M. m. musculus* was performed using a forward primer which binds to exon 2 and a reverse primer that binds to the oligo-dT RACE linker. The PCR reaction was loaded on an agarose gel that is shown in Figure 9. This non-quantitative method reveals that at least a very small amount of exon 2 is present in *M. m. musculus*. This might mean that the musculus promoter has a very weak and basal activity in musculus as well.

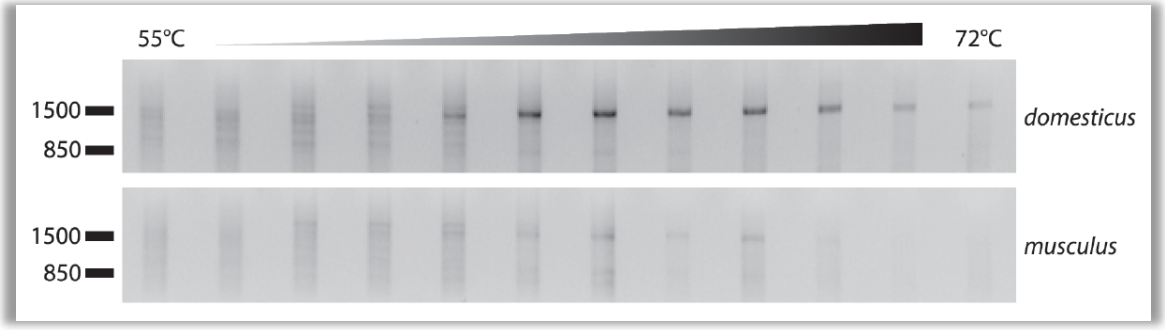


Figure 9 Gradient PCR of *M . m. domesticus* (top) and *M. m. musculus* (bottom) 3' RACE template. Gradient ranges from 55°C to 72°C annealing temperature (increasing from left to right). The forward primer is placed in the Mkk7 alpha specific exon 2. The reverse primer binds to the oligo-dT RACE-linker. Size marker bands indicate kb.

In summary, it can be stated that the Mkk7α1 is a testis specific isoform predominantly expressed in *M. m. domesticus*, but not expressed in *M. m. musculus*.

2.4 Sequence polymorphism in Mkk7 exon 13

The RACE experiments showed that all transcripts of Mkk7 variant 2 end at the same nucleotide position. However, the transcripts of the *M. m. domesticus* specific Mkk7 variant 1 do not. The positions of polyadenylation cleavage sites vary within a stretch of 21 bases. The genomic region encompassing 215 bases upstream and 137 downstream of this stretch was sequenced in 12 animals of *M. m. domesticus* and *M. m. musculus* respectively. The consensus sequences were aligned against a Black6 database sequence (www.ENSEMBL.org) and are presented in Figure 10. As expected, the Black6 sequence is identical to *M. m. domesticus*. A common polyadenylation signal motive (A(A/T)UAAA) is not found in the proximity of the polyadenylation cleavage region, and therefore, a non-canonical mechanism must be used.

Polyadenylation is a widespread and complex process involving the interaction of numerous cis- and trans-acting factors (Graber et al., 1999; Venkataraman et al., 2005). Alternative polyadenylation signals are used in a cell type and developmental stage dependent manner. Particularly during mammalian spermatogenesis there is an enhanced usage of non-canonical polyadenylation that leads to shortened 3' untranslated regions (Liu et al., 2007). The most prominent polyadenylation signals are located within ~40 nucleotides upstream of the cleavage site, but other co-acting cis-elements can be ~100 nucleotides away on both sides (Hu et al., 2005; Lee et al., 2007; Venkataraman et al., 2005).

Two fixed SNPs between *M. m. domesticus* and *M. m. musculus* can be found 66 and 126 nucleotides upstream of the polyadenylation cleavage site. The SNPs are not embedded in any obvious polyadenylation signal. It is not possible to state whether these SNPs are capable of initiating the early termination in the *M. m. domesticus* lineage because too little is known about non-canonical polyadenylation mechanisms. Therefore, it remains unclear whether variant 1 occurs in *M. m. domesticus* due to different cis- or to different trans-factors.

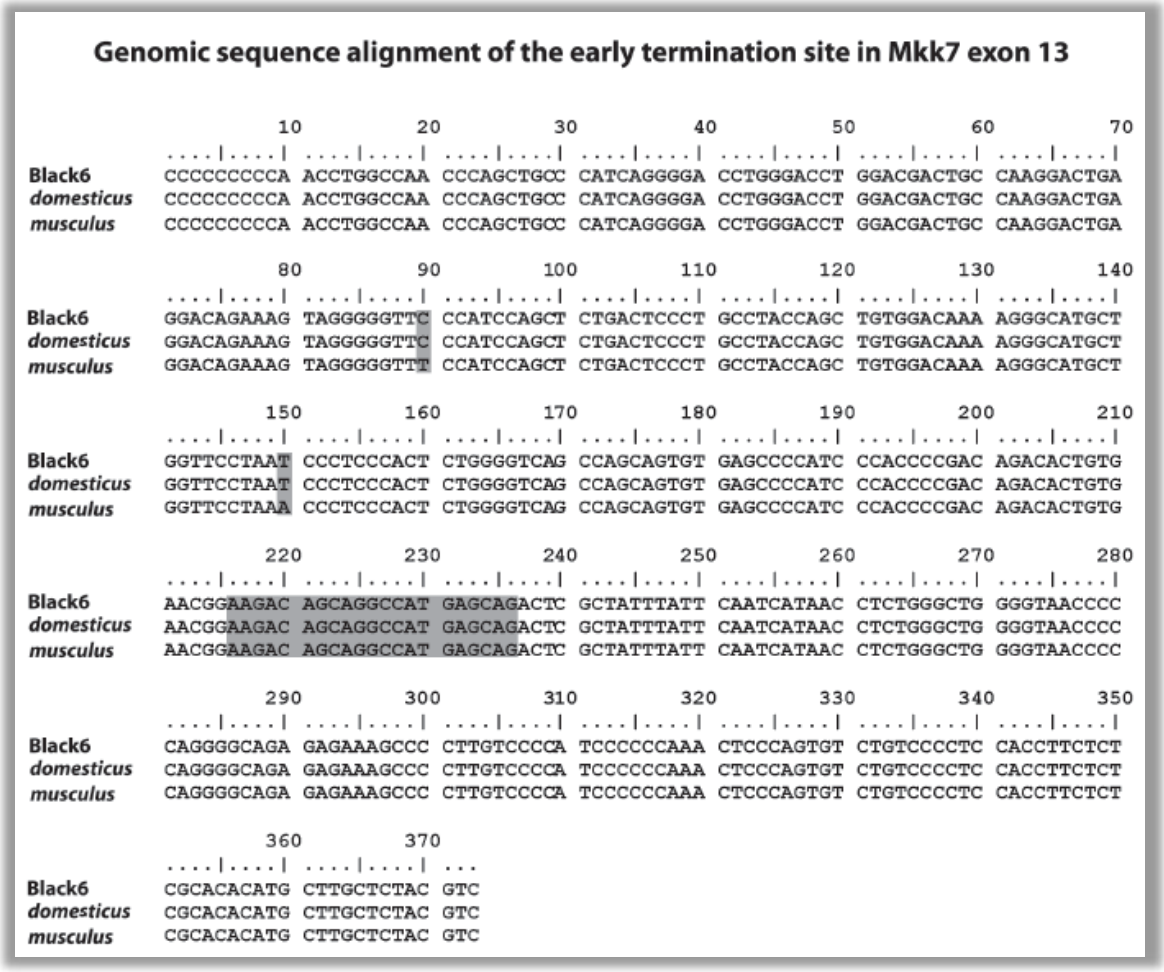


Figure 10 Genomic sequence alignment of a proximal part of Mkk7 exon13 including the transcription terminus of Mkk7*α*1. Consensus sequences of 12 animals each from *M. m. musculus* and *M. m. domesticus* are aligned against Black6 database sequence. Relative base pair positions are indicated. SNP positions are highlighted with grey background color. The Mkk7*α*1 terminus ranges from position 216 to position 236 (marked with grey box).

2.5 Characteristics of the Mkk7 alpha-promoter

2.5.1 Expression pattern of Mkk7 alpha in testis

Testis tissue mainly consists of seminiferous tubules, which are the location of spermatogenesis. Spermatogonial stem cells adjacent to the inner tubule wall divide and form spermatocytes which undergo meiosis. After meiosis the spermatocytes develop into spermatids and change morphologically from round spermatids to elongated spermatids before the generation of mature spermatozoa is completed. The three main

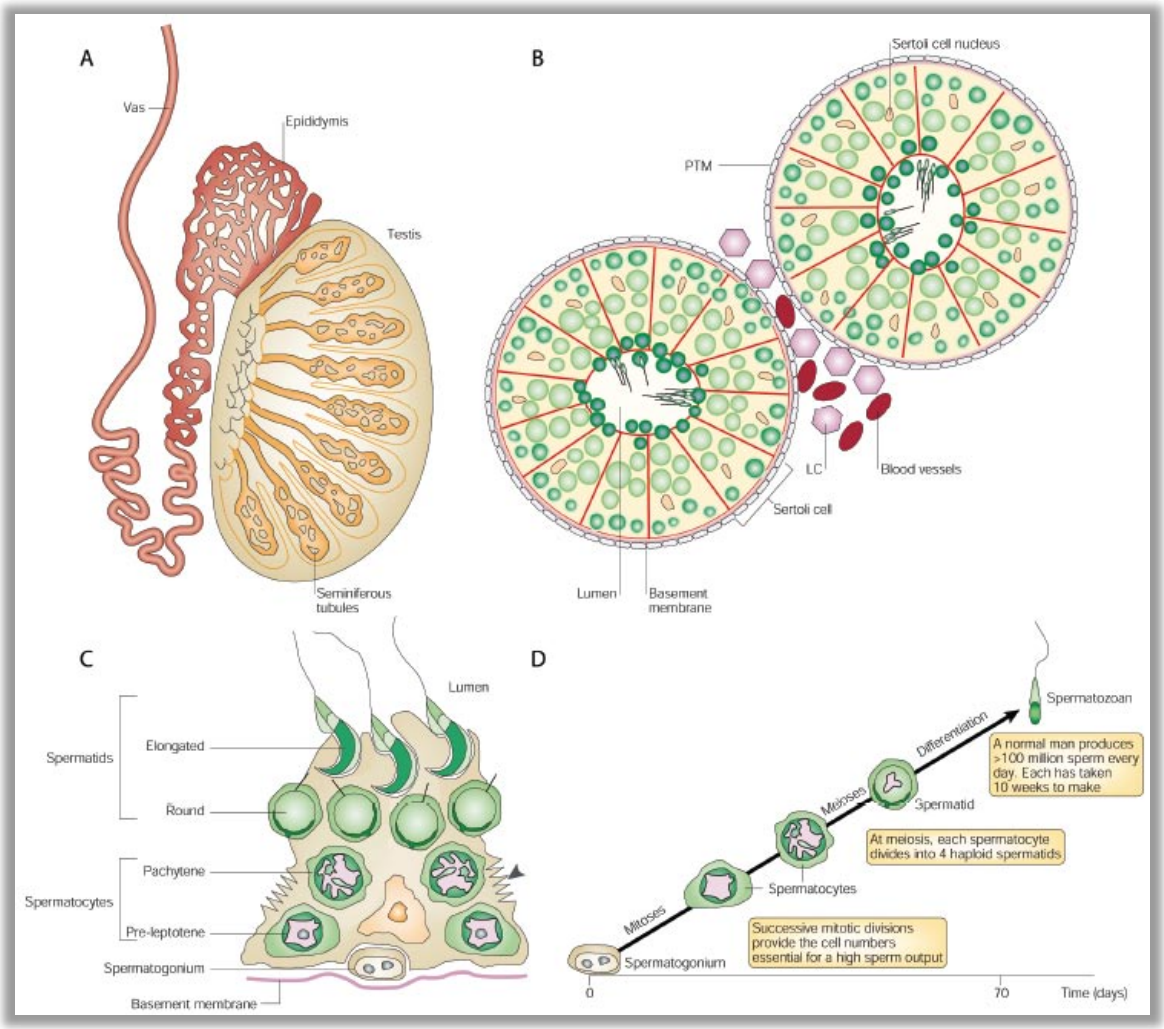


Figure 11 Scheme of spermatogenesis. A) Anatomy of testis and epididymis B) Cross section of seminiferous tubules C) Gametogenic epithelium D) Stages of spermatogenesis. Illustration is taken from (Cooke and Saunders, 2002)).

stages, spermatogonia, spermatocytes and spermatids, are classified into further sub-stages (Russell et al., 1990). Sperm precursor cells are embedded in sertoli cells which define the shape of the spermatogenic epithelium and support the germ cells. Through the influence of sertoli cells, developing sperm precursor cells proceed towards the lumen of seminiferous tubules according to their degree of maturation. Terminal spermiation releases the sperm cells into the luminar fluid of the tubules that transfers them to the epididymis. Hence, ring-shaped zones representing different cell stages can

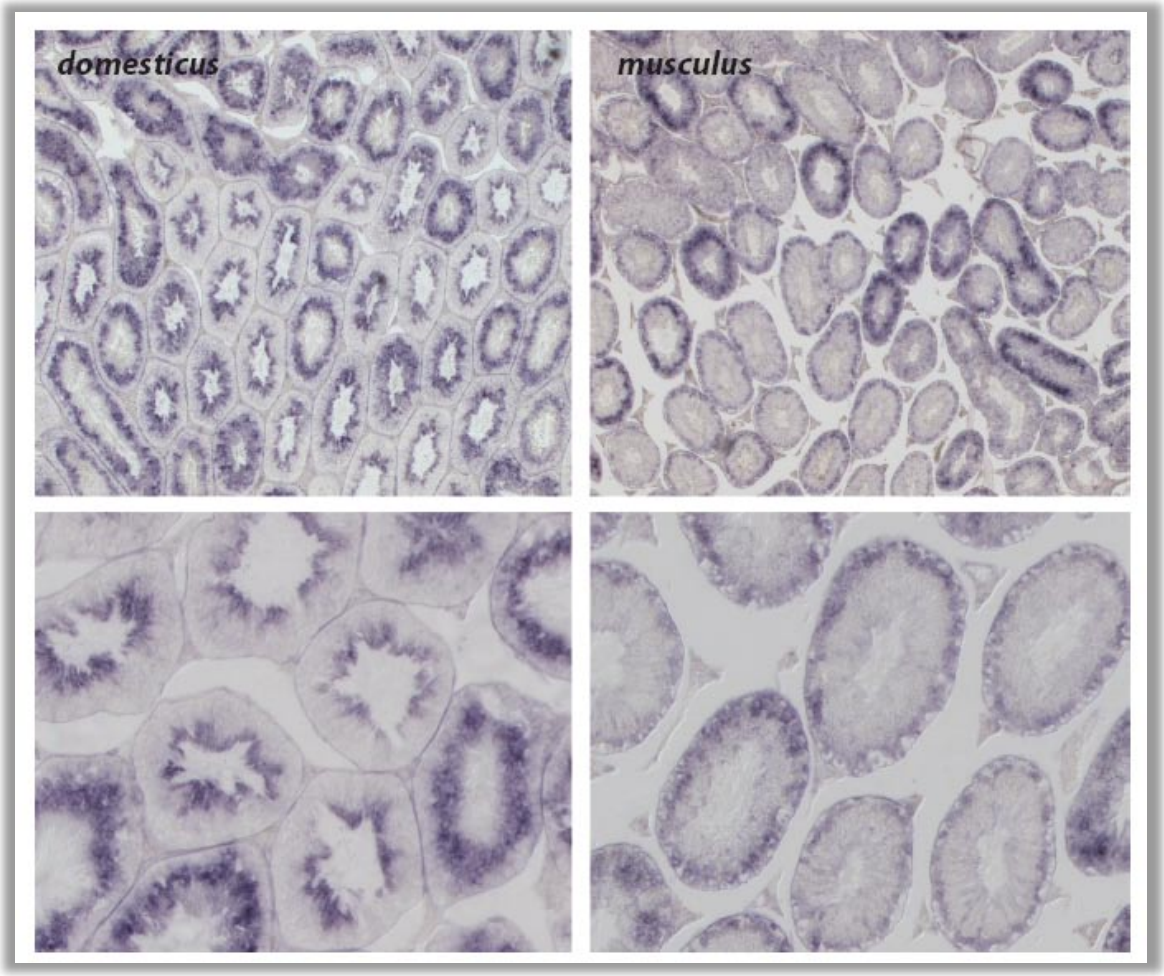


Figure 12 In situ hybridization of a Mkk7 probe to testis sections of *M. m. domesticus* (left) and *M. m. musculus* (right). Pictures were taken at 50 fold magnification (upper row) and 100 fold magnification (lower row).

be distinguished in a transverse section of seminiferous tubules. At the outer ring small spermatogonial cells can sporadically be seen attached to the basal membrane. A medial zone contains spermatocytes and early round spermatids. The inner ring close to the tubular lumen comprises late round spermatids, elongated spermatids and spermatozoa. Spermatogenesis is schematically illustrated in Figure 11.

In situ hybridization on testis sections of *M. m. musculus* and *M. m. domesticus* was performed in order to investigate the localization of Mkk7 mRNA in testicular cells (Figure 12). A strong staining of spermatids can be observed in *M. m. domesticus*. The

detailed view at higher magnification shows that the staining ranges from late round spermatids to elongated stages in the inner circle. In contrast, Mkk7 expression pattern in *M. m. musculus* is restricted to earlier stages (Figure 12). Note that the signal in *M. m. musculus* needed at least four days of incubation before it was clearly visible, while the signal in *M. m. domesticus* appeared within a few hours. Incubation of *M. m. domesticus* sections for 4 days lead to strong over-staining and tissue corruption due to substrate precipitation. So, it was not possible to rule out if Mkk7 mRNA is also present in earlier stages as found in *M. m. musculus*. The strong expression difference between *M. m. domesticus* and *M. m. musculus* is due to the alpha promoter. Therefore the fact that the spermatid specific staining in *M. m. domesticus* arises quickly, while it is still absent in *M. m. musculus* after much longer staining, implies that the alpha promoter is spermatid specific. Moreover, it can be stated that Mkk7 is not expressed in late spermatids in the *M. m. musculus* lineage.

2.5.2 Sequence polymorphism in the Mkk7 alpha promoter region

As mentioned above, it is known that the expression difference between both subspecies derives from cis-effects (Harr et al., 2006). Therefore, it is worthwhile to look at sequence polymorphism between *M. m. domesticus* and *M. m. musculus* in order to search for candidate mutations that might be responsible for the enhancement of Mkk7 alpha expression in *M. m. domesticus*. According to the transcript structure, the basal alpha promoter must be located in intron 1. Therefore, the Mkk7 intron 1 and flanking regions were sequenced in Black6, *M. m. domesticus* and *M. m. musculus*. Consensus sequences of 8 animals per subspecies were generated and aligned with Black6 (Supplement A). The result obtained from the sequence data is summarized in Table 2. Assuming that the expression difference between *M. m. musculus* and *M. m. domesticus* was the target for natural selection, it can be expected that the mutation which caused the effect is fixed between the populations compared. Five fixed differences between *M. m. musculus* and *M. m. domesticus* can be found, but only one fixed SNP is present in a 500 bp window upstream of the transcription start. Empirical data provide evidence that specific transcription activity in reproductive tissues and particularly in spermatogenesis

Polymorphism in Mkk7 intron 1

Position of SNP relative to transcription start	Type of SNP	Allele frequencies in *M. m. domesticus*				Allele frequencies in *M. m. musculus*			
-12	Transition	A	0	G	1	A	0.0625	G	0.9375
-84	Transversion	A	1	T	0	A	0	T	1
-297	Transition	T	0	C	1	T	0.0625	C	0.9375
-466	Transversion	G	0.125	T	0.875	G	0	T	1
-488	Indel	A	0.5625	-	0.4375	A	0	-	1
-502	Transversion	C	0	A	1	C	1	A	0
-702	Transversion	T	0.25	A	0.75	T	1	A	0
-734	Transversion	C	0	G	1	C	1	G	0
-762	Transition	C	0	T	1	C	1	T	0
-1035	Transition	A	0.125	G	0.875	A	0	G	1
-1052	Transition	A	0	G	1	A	1	G	0
-1087	Transversion	G	0.125	T	0.875	G	0	T	1

Table 2 Frequencies of SNPs in *M. m. musculus* and *M. m. domesticus*, type of SNPs and their positions relative to the transcription start of Mkk7a1. SNPs that are fixed in both subspecies are highlighted with grey background.

is regulated by very short proximal promoters (Blaise et al., 2001; Han et al., 2004; Li et al., 1998; Reddi et al., 1999; Scieglinska et al., 2004; Topaloglu et al., 2001; Zambrowicz et al., 1993). Those studies demonstrated that proximal promoters shorter than 300bp, or even less than 100bp, are sufficient to drive spermatid specific expression in mice. Unique mechanisms of gene regulation are postulated to exist in post-meiotic cells (Acharya et al., 2006; Somboonthum et al., 2005). Additionally, it was recently shown that a 5'-ACACAC motive ~170 bp upstream of the transcription

start serves as in insulator in the spermatid specific expression of the mouse SP-10 gene and it was suggested that insulators might generally play an important role in maintaining spermatid specific transcription (Abhyankar et al., 2007; Acharya et al., 2006; Reddi et al., 2003; Reddi et al., 2007).

Based on these facts, it is possible that the fixed SNP at -84 bp relative to the Mkk7 alpha transcription start might raise the expression level in *M. m. domesticus*. A 5'-ACACAC motif can be found at position -259 in the Mkk7 alpha promoter. The following sections address the question of whether the mutation at -84 and the 5'-ACACAC motive at -259 influences the cis-regulatory properties of the Mkk7 alpha promoter. Therefore, a suitable reporter gene system was established to test this.

2.5.3 Establishing a reporter system for in vivo applications

To investigate cis-regulatory differences such as the Mkk7 example it is necessary to use reliable and sensitive reporter assays to test single mutations that may influence gene expression. Future studies will probably raise more such examples of differentially expressed adaptive trait genes in mice that will be candidates for follow-up experiments. A reporter system for those genes has to fulfill certain requirements. It must offer the potential for precise measurement of expression levels and expression pattern. It should be applicable in vitro and in vivo. Most promoters can easily be tested in appropriate cell culture models, but some tissue specific promoters might not, because appropriate cell culture models are difficult or not available. A suitable reporter system was chosen and developed with regard to its application in future studies. The results are presented in the following.

Luciferases offer very good performance for the quantification of gene expression and they are commonly used for that purpose (Sadikot and Blackwell, 2005). The bioluminescence of luciferases can precisely be measured as a rate of expression strength after substrate application. The most popular luciferases are the firefly luciferase (from *Photinus pyralis*) and renilla luciferase (from *Renilla reniformis*) (Zinn

et al., 2008). Both luciferases catalyze different substrates, namely luciferin for firefly and coelenterazine for renilla, and can be used for parallel measurements in cell culture experiments and for live imaging in vivo (Bhaumik and Gambhir, 2002; Perkinton et al., 2004; Ray and Gambhir, 2007). It was reported that hrLuc, a novel synthetic variant of renilla luciferase, is more sensitive than other common luciferases. The recent invention of new substrates for in vivo application of renilla luciferases improved their application spectrum (Otto-Duessel et al., 2006). Thus, hrLuc was chosen as the reporter gene in this study.

Luciferase is an optimal tool for the quantification of gene expression, but it does not allow precise distribution of gene expression patterns in tissues. Evolutionary changes of cis regulatory regions might affect the expression strength and the expression pattern of genes. Therefore, a MYC-tag (EQKLISEEDL) (Evan et al., 1985; Hilpert et al., 2001) was fused to the C-terminus of hrLuc. Antibody staining can detect this tag within tissues (Hald et al., 2003) and help to investigate the cell-type specificity of the analyzed promoter. The resulting reporter was named Luc-MYC. Luc-MYC and hrLuc were cloned into an expression plasmid under the control of a CMV-promoter. The activity of Luc-MYC was compared to normal hrLuc in cultured C2C12 myoblast cells (Yaffe and Saxel, 1977). Figure 13 shows that the MYC tag does not negatively affect the luciferase activity. Therefore Luc-MYC was used in the following experiments.

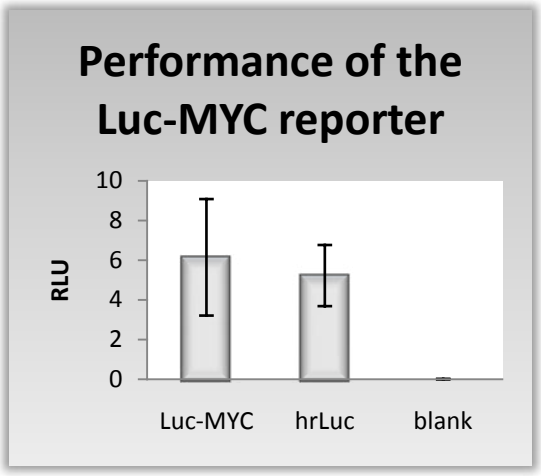


Figure 13 Activity of the hrLuc in comparison to the MYC-tagged variant Luc-MYC. Both reporter genes were cloned into a mammalian expression plasmid under the control of CMV promoter and transfected into C2C12 myoblast cells. Relative light units (RLU) were measured after the application of substrate. Luc-MYC represents the median of 6 independent measurements, hrLuc and blank (no vector) represent both the median of 3 independent transfections. Error bars represent standard deviations.

Reporter gene expression in cell culture systems is an easy and commonly used method (Birgersdotter et al., 2005; Shiffman and Porter, 2000). The expression of reporter genes generated through transgenesis for imaging in living mice is far more sophisticated. A common technique to express reporter genes in vivo is the generation of conventional transgenic mice by pronucleus injection (Gordon et al., 1980; Okabe, 2001; Saunders, 2003). The disadvantage of conventional transgenic mice is that the reporter genes are present at variable numbers of random integrants. Typing of founder animals is time consuming. Different copy numbers and the local effects of different integration sites may influence the activity of reporter genes. A more defined way to obtain in vivo reporter mice is achieved by gene targeting via homologous recombination in embryonic stem (ES) cells (Doetschman et al., 1987; Thomas and Capecchi, 1987). On the other hand, this method is also very labor intensive. The availability of ES cells from *M. m. domesticus* and *M. m. musculus* wild or wild derived mouse strains could facilitate the comparison of natural promoter variants by gene targeting, but such cell lines do not exist yet. In recent years, bacterial artificial chromosome (BAC) targeting increased in importance as a new technique in mouse transgenesis (Yang et al., 1997). BACs containing the gene of interest can be modified by site directed recombination in *Escherichia coli* (Cotta-de-Almeida et al., 2003; Muyrers et al., 1999). Thereby, reporter genes can be inserted into the target locus. The modified BACs can be used as large targeting vectors in highly efficient ES cell transfections and save time by avoiding molecular cloning of targeting vectors (Valenzuela et al., 2003). In most cases, Black6 BAC libraries (Osoegawa et al., 2000) can serve as a source for *M. m. domesticus* DNA and a *M. m. musculus* derived PWD-strain BAC library is available (Jansa et al., 2005), but screening of BAC libraries is time consuming and has to be repeated in every new experiment. Moreover, the screening and characterization of recombinants is costly and time consuming. An alternative approach is site directed mutagenesis in ES cells mediated by recombinases (Araki et al., 1997). Therefore, established mouse ES cell lines that contain a recombinase recognition site are co-transfected with a vector that contains the same site and another vector expressing the respective recombinase. The recombinating vectors can be used as shuttles to integrate

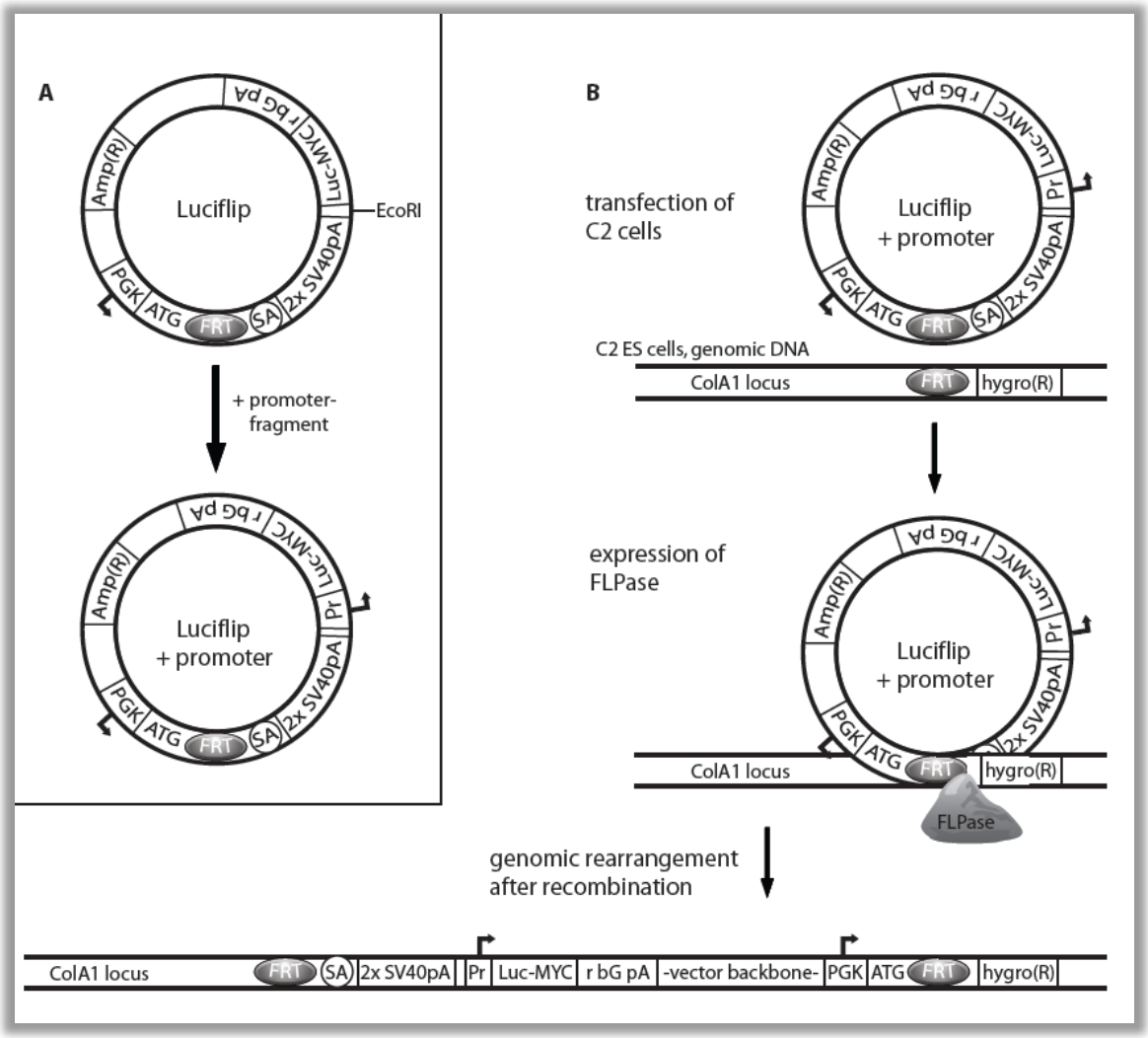


Figure 14 Description of the Luciflip reporter plasmid. (A) Promoter fragments can be inserted upstream of the renilla-MYC-tag luciferase (Luc-MYC) by a single cloning step using the unique EcoRI site. A 3' intron and a polyA signal from the rabbit beta globin gene (r bG pA) are placed downstream of the Luc-MYC coding sequence in order to enhance the translation efficiency of the gene product. A cassette containing a splice acceptor site (SA) followed by two SV40 polyA signals (2x SV40pA) avoids disturbance potentially caused by upstream cis effects. (B) The plasmid Luciflip can be used to generate single copy reporter insertions in combination with the C2 mouse embryonic stem cell line. C2 cells contain a hygromycin resistance cassette lacking a promoter and a functional start codon (hygro(R)) downstream of the collagen type I alpha 1 locus (ColA1). FRT sites in the genomic locus and in the Luciflip plasmid enable site directed recombination of the Luciflip vector DNA with the genome after co-transfection of C2 cells with Luciflip and a FLPase expression vector leading to the insertion of the reporter construct. After recombination, a PGK promoter and a start codon (PGK ATG) complete the hygromycin resistance cassette allowing selection for positive clones.

DNA sequences of choice into the genome of ES cells. Such systems are available using the Cre/loxP system and the FLPase/FRT system (Beard et al., 2006; Masui et al., 2005). The C2 and KH2 ES cell lines have a FRT site inserted downstream of the collagen A1 locus (Beard et al., 2006). Co-transfection of these ES cells with a FLPase expression plasmid and an FRT containing shuttle vector leads to FLPase mediated site directed integration of the shuttle vector downstream of the collagen A1 locus. This "flip-in" procedure is highly efficient and it was already shown that it is suitable for the defined expression of genes in vivo (Beard et al., 2006). C2 and KH2 ES cell lines are derived from F1 hybrids of different mouse strains. Therefore, they can be used for microinjection into tetraploid blastocysts. Animals that are raised by this technique are not chimeric and can directly be analyzed or mated for later experiments (Nagy et al., 1993). This FRT/FLPase system was chosen to be modified for fast luciferase reporter construction, because of its high efficiency and its methodical benefits.

The pBS31' plasmid (Beard et al., 2006) was modified and the resulting vector was named Luciflip. The full vector sequence of Luciflip is available in the digital supplement. In order to test cis-regulatory sequences, promoter fragments can be cloned into a solitaire EcoRI restriction site upstream of the Luc-MYC coding region (Figure 14A). A cassette consisting of a splice acceptor and two SV40 polyA signals protects the reporter expression from transcriptional upstream activity. An intron stabilizes the expression of Luc-MYC. Luciflip can serve as shuttle vector for site directed insertions in combination with C2 or KH2 ES cells (Figure 14B). To control for the functionality of Luciflip, a CMV promoter was cloned into the EcoRI site. NIH/3T3 fibroblast cells (Todaro and Green, 1963) were transfected with the Luciflip-CMV construct and with empty Luciflip plasmid. The CMV promoter provoked a very high luciferase activity while the pure Luciflip plasmid without promoter gave only low background activity (Figure 15). This result demonstrates that Luciflip can also be used to quantify expression strength in common cell culture experiments.

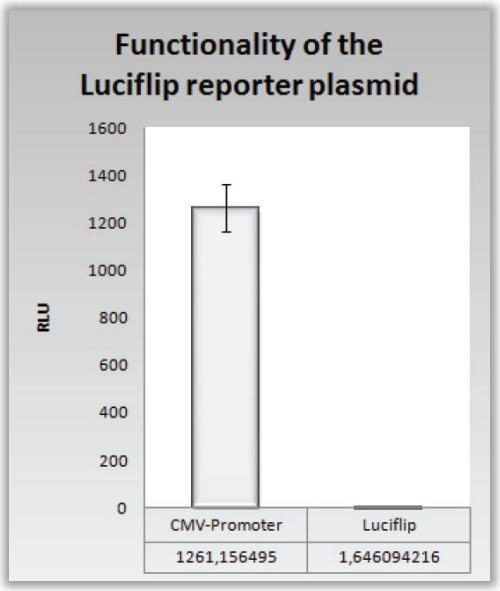


Figure 15 Performance of the Luciflip reporter plasmid. NIH/3T3 fibroblasts were transfected with Luciflip plasmid containing a CMV promoter or no promoter. Relative light units (RLU) were measured after substrate application. The signal of a co-transfected firefly luciferase was used to normalize for transfection efficiency. The values display the median of 8 independent transfections each. Error bars indicate standard deviations.

2.5.4 Characterization of the Mkk7 alpha promoter by Luciferase reporter Assay

The question of whether the transversion at -84 and the 5'-ACACAC motive at -259 influences the cis-regulatory properties of the Mkk7 alpha promoter was attempted to be answered with the help of a luciferase expression assay in cultured cells. The Mkk7 alpha expression which is of most interest in this context is restricted to late spermatids. Culturing this type of cells is very difficult due to its haploid post meiotic stage with condensed chromatin. An established spermatid cell culture model was not available. Alternative cell lines have the disadvantage that they will most likely not recognize the Mkk7 alpha promoter. In the absence of better options, the NIH/3T3 fibroblast cell line was chosen for the experiment.

Relative to the transcription start, -487/+43 wild type fragments of the Mkk7 alpha promoter were generated by PCR. The only fixed difference between *M. m. domesticus* and *M. m. musculus* within this region is the transversion at position -84. It cannot be expected that this fragment is sufficient to drive luciferase expression in non-spermatid cells, but it is likely that the expression level can be raised by deleting the 5'-ACACAC motive at -259, if the assumption is correct, that this sequence maintains spermatid

specific transcription by acting as an insulator in other cells. Thus, -487/+43 fragments lacking the 5'-ACACAC motive at -259 were generated as well. It can be assumed that the -487/+43 fragments do not contain enhancer elements which promote expression in fibroblasts. Therefore, a CMV enhancer was ligated upstream to both versions. All four constructs (wild type, deleted insulator, CMV enhancer + wild type, CMV enhancer + deleted insulator) were created as *M. m. domesticus* variants with an adenine at position -84 and as *M. m. musculus* variants with a thymine at position -84. The 8 different constructs were cloned into the Luciflip plasmid. NIH/3T3 cells were transfected with the reporter constructs in 8 technical replicates of each. A firefly luciferase expression plasmid was co-transfected to normalize the renilla signals. Reporter activity was measured after substrate application.

Figure 16 summarizes the result of the reporter assay. The *M. m. domesticus* variant generates significantly higher signals compared to *M. m. musculus* in every scenario.

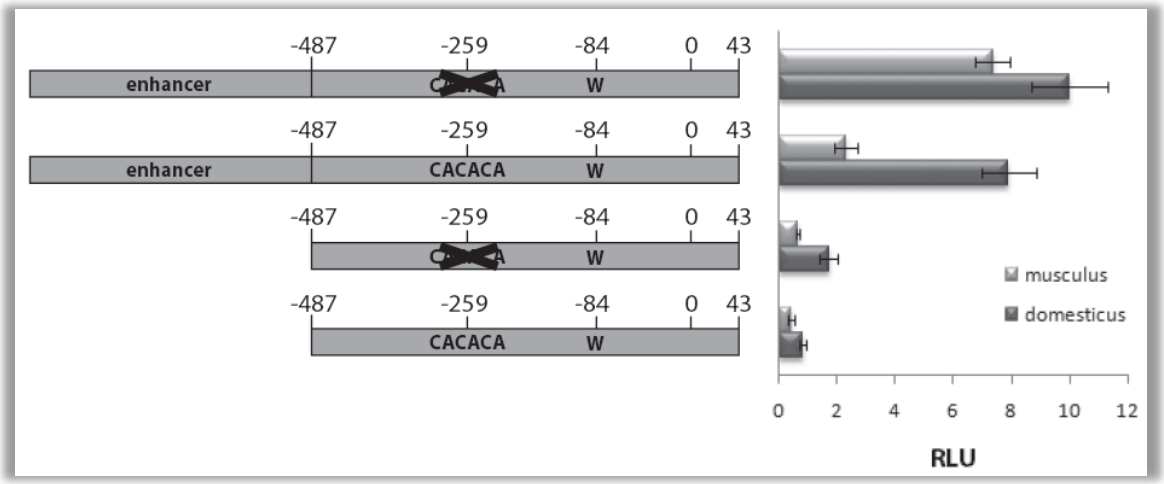


Figure 16 Reporter assay with different variants of Mkk7 alpha promoter. *M. m. musculus* and *M. m. domesticus* promoter fragments (-487/43) were tested as wild type, with deleted insulator, as wild type and upstream enhancer and with deleted insulator and upstream enhancer. Each fragment was co-transfected together with a firefly luciferase expression vector into NIH/3T3 fibroblast cells in 8 independent replicates. Luciferase activity was measured after substrate application. The renilla signals were normalized with the firefly signals to eliminate disturbing variation caused by different transfection efficiencies. The values represent the median of the 8 replicated transfections for every construct. Error bars indicate standard deviations.

The different replicates are consistent, indicated by relatively small standard deviations. Deletion of the 5'-ACACAC motive at -259 indeed increases the expression strength in all variants. The presence of an enhancer potentiates the effects as expected.

These data provides strong evidence that the adenine at position -84 enhances the activity of the basal promoter. For this reason it can be supposed that a contribution of this mutation to the expression difference between *M. m. domesticus* and *M. m. musculus* in late spermatids is very likely. The sequence 5'-ACACAC represses the action of an adjacent enhancer to a certain extent. This finding supports the hypothesis that it acts as an insulator in the Mkk7 alpha promoter.

2.6 Relation of Mkk7 expression to protein status in testis

The role of Mkk7 in the MAP signaling cascades is described in section 2.1. So far, the only known function of Mkk7 is the activation of JNK. Western blotting with antibodies against JNK and phosphorylated JNK was performed in order to evaluate the contribution of JNK activation in testis (Figure 17). High amounts of total testis protein from *M. m. musculus* and *M. m. domesticus* were blotted and anti-phospho-JNK antibody was applied for detection. Cell extract of UV-treated 293 cells (Graham et al., 1977) served as a positive control for JNK activation. Untreated 293 cell extract

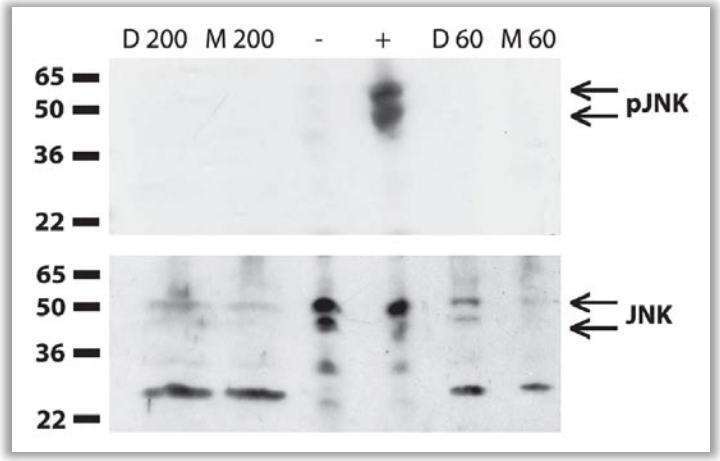


Figure 17 JNK Western blot. 200 µg and 60 µg testis protein of *M. m. musculus* (M 200, M 60) and of *M. m. domesticus* (D 200, D 60) was blotted and stained with anti-phospho-JNK antibody. Extracts of UV-treated 293 cells (+) and untreated 293 cells (-) served as controls. The blot was stripped and re-probed with an anti total JNK antibody. Protein ladder bands indicate kDa. Arrows indicate the positions of phosphorylated JNK (pJNK) and total JNK (JNK).

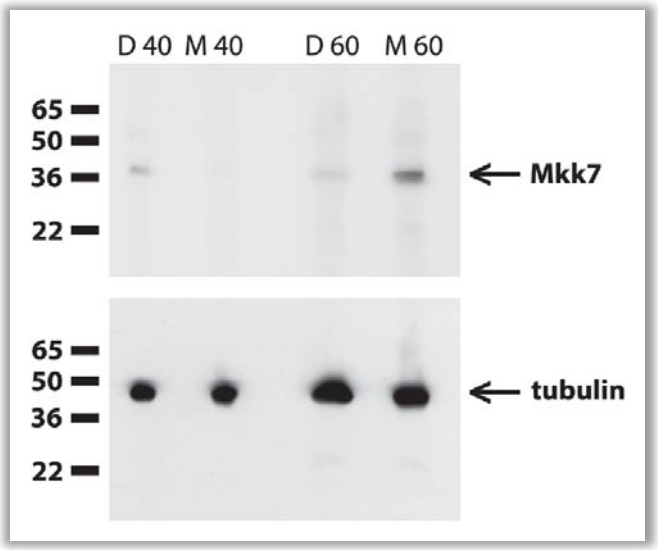


Figure 18 Mkk7 Western blot. 40 µg and 60 µg testis protein of *M. m. musculus* (M 40, M 60) and of *M. m. domesticus* (D 40, D 60) was blotted and stained with anti-Mkk7 antibody. The blot was stripped and re-probed with an anti-tubulin antibody. Protein ladder bands indicate kDa. Arrows indicate the positions of Mkk7 and tubulin.

was used as negative control. No phosphorylated JNK protein was detected in the testis samples and in the untreated 293 cell extracts, while the UV-treated cells triggered a clear staining. As a control, the same blot was re-probed with an anti-total-JNK antibody. Total JNK was detected in all lanes although the concentration of JNK in testis is very low despite the huge amount of testis protein that was loaded on the gel. On that account, it can be reasoned that the JNK pathway does not have a consistent function in spermatogenesis.

Detection of Mkk7 protein in testis by Western blot yielded an unclear picture (Figure 18). Two different concentrations of total testis protein from *M. m. musculus* and *M. m. domesticus* (40µg and 60µg each) gave two different outcomes. The *M. m. domesticus* sample generates a higher Mkk7 signal in comparison to *M. m. musculus* at low concentrations, but the situation is different at higher concentration. Re-probing the blot with anti-tubulin antibody showed that the amount of loaded protein in the *M. m. musculus* - *M. m. domesticus* comparisons is equal. This paradoxical observation could be reproduced in several independent experiments (Daniela Häming, personal communication). However, if the protein levels reflect the RNA levels, a striking disparity in favor of *M. m. domesticus* would be expected. This can be excluded according to the Mkk7 Western blot.

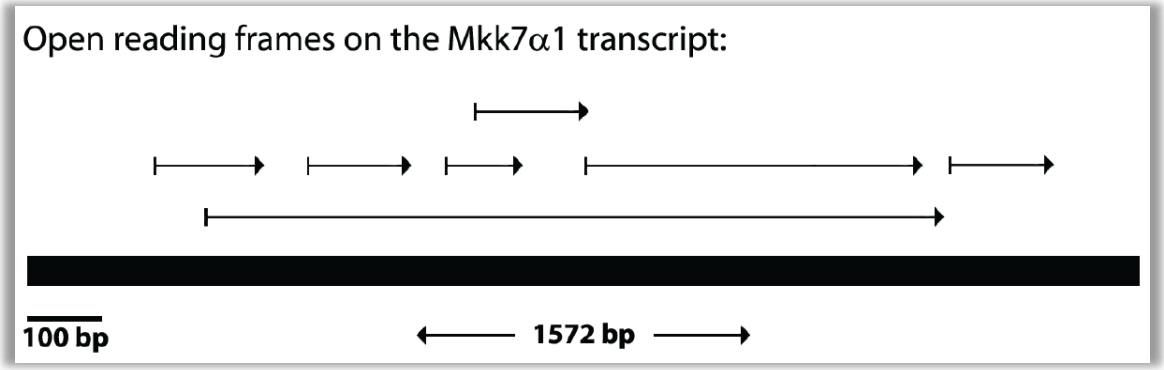


Figure 19 Open reading frames on the Mkk7α1 RNA.

It can be asked why the Mkk7α1 transcript evolved in domesticus to be strongly expressed in spermatids although the protein level is not much affected. In a previous study, the Mkk7α1 isoform was postulated to be a kinase variant with truncated N-terminus (Tournier et al., 1999). Within the framework of this article, Mkk7α1 protein was expressed in bacteria. Biochemical analyses showed that it possesses very little capability to activate JNK unlike all of the other isoforms. This finding is not surprising, because the missing part of the N-terminus contains docking sites that mediate the interaction with JNK (Ho et al., 2006). Mkk7α1 is a spermatid specific transcript and the JNK pathway does not seem to play a role in spermatogenesis. Therefore, a so far unknown role of Mkk7 in spermatogenesis can be inferred.

Analysis of open reading frames (ORF) in the Mkk7α1 RNA raises a hypothesis about the function of this Mkk7 variant. Figure 19 shows all ORFs on Mkk7α1 RNA that start with an ATG codon and are longer than 30 amino acids. The longest ORF (346 aa / 1038 bp) represents the postulated Mkk7α1 isoform. Obviously, another short ORF (50 aa / 150 bp) has an earlier start codon. It is designated ^{sp}Mkk7 (small peptide translated by Mkk7 RNA) and codes for the following peptide:

MGAAAHHPQRAPHSTLHPPPGPATCWGSHQPCSHRAVWRASRLTRSCRRS.

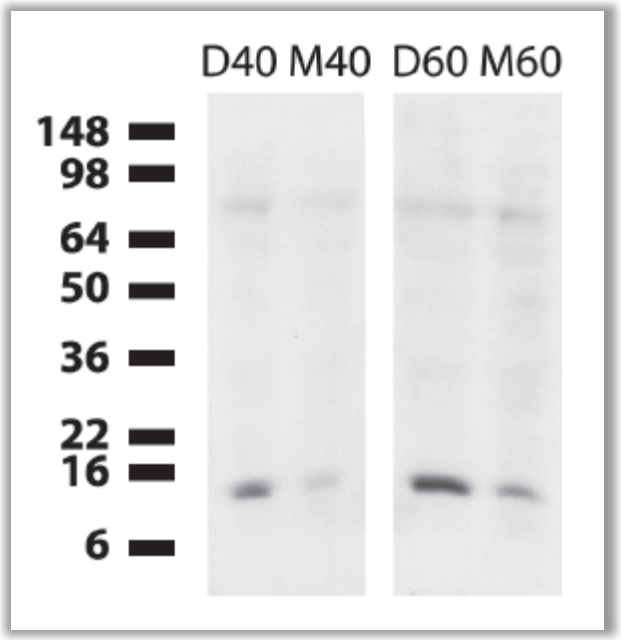


Figure 20 Mkk7 Western blot. 40 µg and 60 µg testis protein of *M. m. musculus* (M40, M60) and of *M. m. domesticus* (D40, D60) was blotted and stained with anti-^{sp}Mkk7 antibody. Protein ladder bands indicate kDa.

The peptide does not match to any known protein and does not contain any known domains (www.expasy.org;http://blast.ncbi.nlm.nih.gov/Blast.cgi). The nucleotides surrounding the start codon of ^{sp}Mkk7 (TGGCCAACG **ATG** G) matches much better to the Kozak-consensus-sequence (Kozak, 1984, 1986, 1987) than the nucleotides surrounding the start codon of Mkk7α1 (CCCCGCCAC **ATG** C). A purine at position -3 and a guanine at position +4 are the most important sequence elements for the initiation of translation. To find out whether ^{sp}Mkk7 is indeed translated, an antibody against an immunogenic stretch of the peptide (AHHPQRAPHSTLHPP) was generated and Western blot with testis protein from *M. m. musculus* and *M. m. domesticus* (40µg and 60µg each) was performed (Figure 20). The calculated mass of the peptide is ~5.5 kDa while the predicted isoelectric point is ~pH12.0 (http://www.expasy.ch/tools/pi_tool.html). Hence, a slower migration in SDS-page due to positive charges is expected. The anti-^{sp}Mkk7 antibody specifically stained a band which matches to this data. This band is significantly stronger in *M. m. domesticus* compared to *M. m. musculus*.

Based on these findings, I propose the following hypothesis. ^{sp}Mkk7 is translated in both sub-species that were analyzed, but Mkk7α1-kinase-protein is not translated at all or very weakly. In *M. m. musculus*, it is translated from Mkk7 beta during early stages of spermatogenesis. In *M. m. domesticus*, the amount ^{sp}Mkk7 peptide is increased and its presence is extended to late spermatids by the expression of Mkk7α1-RNA.

2.7 Summary

Mkk7 shows a different testis expression in comparison between *M. m. domesticus* and *M. m. musculus*. The locus exhibits clear signs of a selective sweep in the *M. m. domesticus* population. Northern blotting, RACE and qRT-PCR experiments identified the short Mkk7α1 isoform being specifically and very highly expressed in *M. m. domesticus*. This isoform is determined by a non-canonical polyadenylation signal and it is expressed by a spermatid specific promoter located in intron1. A luciferase reporter system was established which allows the efficient localization and quantification of reporter gene expression in cell culture and in living mice. This system was used to show that a single nucleotide polymorphism is likely to cause the expression of Mkk7α1 in *M. m. domesticus*. A 5'CACACA motive is responsible for repression of Mkk7α1 expression in non-spermatid cells. The JNK pathway is not active in testis and therefore a so far unknown function of Mkk7 in spermatogenesis was assumed. Mkk7 protein level was not found to be different in *M. m. musculus* and *M. m. domesticus*, but a novel small peptide with higher concentration in *M. m. domesticus* was discovered. It is hypothesized that this peptide is translated from Mkk7α1 RNA. Further it is hypothesized that this peptide has a function in spermatogenesis and was the target of natural selection in *M. m. domesticus*.

3 LP10

3.1 LP10 is a potential adaptive trait gene with unknown function

The gene 1700125F08Rik, designated LP10, was originally found to be differentially expressed in a microarray screen comparing testes of *M. m. domesticus* and *M. m. musculus* (Christian Voolstra and Diethard Tautz, unpublished data). The gene was identified from an EST corresponding to genomic sequence on chromosome 10. Another EST based gene with provisional status, designated Hennes, is transcribed from the opposite strand of the same locus. Both gene names are preliminary working titles that were used in the laboratory. Permanent names are not determined yet, because the function of the genes is still unclear. LP10 ESTs were isolated exclusively from testis, while Hennes ESTs were found in testis and also visual cortex. The mRNA annotation of both genes in the Mouse Genome Informatics database from the Jackson laboratory (http://www.informatics.jax.org/) is shown in Figure 21. The 3' transcript structure of Hennes in testis is not complete, because the only known full length mRNA originated from visual cortex (AK158810). Only a few ESTs from the 5' region have been found in testis. The translational status of Hennes is unclear. LP10 potentially encodes two open reading frames on exon 3 (ORF1: 106 aa, ORF2: 128 aa). The second ORF is predicted to be translated in the official annotation, but without any clear reason. Nothing is known about the function of LP10 and Hennes and similarities in protein sequence of both LP10 ORFs with other known proteins do not exist.

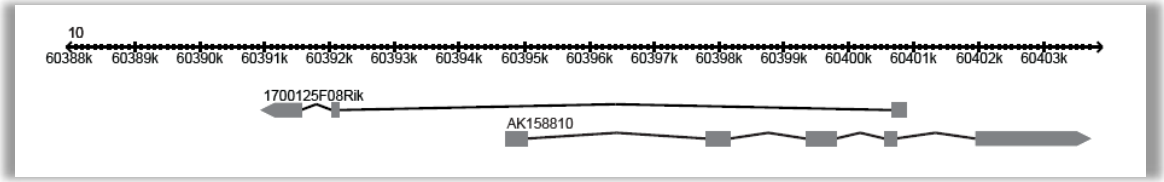


Figure 21 Exons of the mouse LP10/Hennes locus. Annotations of mRNAs are shown in the genomic context of chromosome 10. Source: http://www.informatics.jax.org/, 2008

In the course of a study devoted to population genetics it was found that the Lp10/Hennes locus shows a significant signature of selection (Fabian Staubach and Diethard Tautz, unpublished data): Tajima's D and nucleotide diversity (π) are decreased in the *M. m. musculus* group compared with *M. m. domesticus*. The data is based on two *M. m. musculus* populations (one from the area of the Kazakh city Almaty and another one from the Czech Republic) and on two *M. m. domesticus* populations (from Germany and France). The Lp10/Hennes locus is the only transcribed region in a 250 kb window. The flanking genes are situated in a distance of 100 kb and 150 kb respectively and the reduced variation in the *M. m. musculus* populations peaks at the position of Lp10/Hennes. These findings make the locus an interesting candidate for follow up analysis.

The chromosomal region containing LP10/Hennes is syntenic with homologous regions of rat chromosome 20 and human chromosome 10. The neighboring genes Unc5b and Pcbd1 can be found in the same order in rat and humans and the genomic sequence of LP10 and Hennes can be aligned with the rat and human genome. Nevertheless, no orthologous genes of LP10 and Hennes have yet been annotated in other species. Therefore it can be asked, if LP10 and Hennes are orphan genes that acquired transcriptional activity exclusively in the mouse lineage by de novo mutations.

To further investigate the function of LP10/Hennes in the mouse, a knock-out line was established by gene targeting (Thomas and Capecchi, 1987). It is possible that these genes could be involved in spermatogenesis, because LP10 is exclusively expressed in testis and Hennes expression is limited to testis and visual cortex. Therefore, the knock-out mice were screened for phenotypic abnormalities of the male reproductive system. Gene expression in testes of knock-out mice was compared to expression in wild type mice using microarray analysis to determine whether LP10 and Hennes have influence on the transcriptional activity of other genes. Finally, Northern blot experiments were performed to find out whether LP10 and Hennes are expressed in different species of the Murinae family.

3.2 Targeting strategy

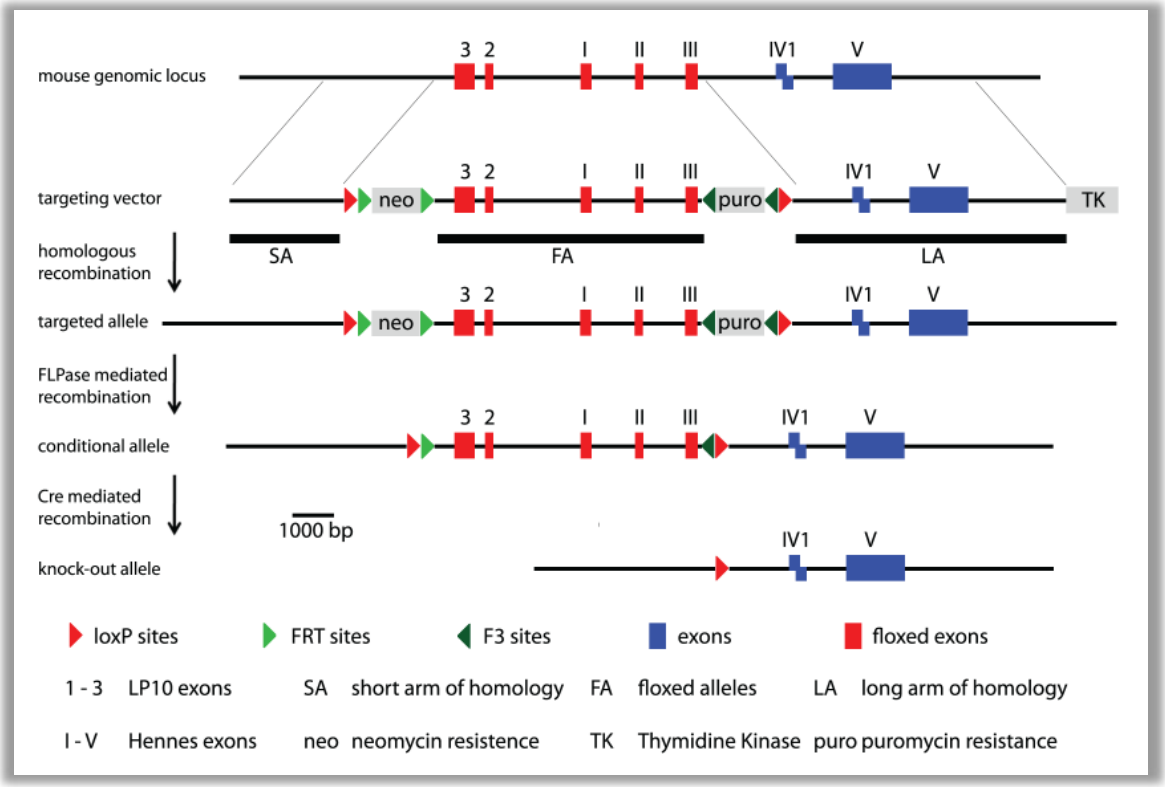


Figure 22 Strategy of LP10/Hennes conditional knock-out. A 7887 bp fragment was flanked by loxP sites. One loxP site was placed downstream of the last LP10 exon. The second loxP site was placed into the first intron of LP10 without affecting exons of Hennes and its promoter. The targeting vector contained an FRT-flanked neomycin resistance cassette and a F3-flanked puromycin cassette in order to increase the co-recombination frequency of both loxP sites after selection. Thymidine kinase served as negative selection marker to decrease the number of random integration events. The short arm of homology was 3282 bp in length and the long arm of homology had a size of 7941 bp. The neomycin and puromycin resistance cassettes were removed by transient transfection of homologous recombinants with FLPase. One FRT and one F3 site each remained in the genome. Cre mediated recombination leads to deletion of LP10 exons 1-3 and to a total inactivation of Hennes due to promoter disruption.

A knock-out of LP10 and Hennes was aimed at investigating their potential function in testis. However, it was anticipated that a total knock-out of LP10 and Hennes could result in sterility or possibly even embryonic lethality. Therefore, it was decided to generate a conditional double knock-out of LP10 and Hennes. A common strategy for generating conditional knock-out mice is achieved by homologous recombination in

embryonic stem cells in combination with the Cre/loxP system (Gu et al., 1993; Rajewsky et al., 1996). The genome of mice carrying conditional alleles is manipulated in such a way that loxP sites flank important parts of the gene of interest without disturbing its function. Crossing those animals with Cre expressing mice results in the Cre mediated deletion of the sequence between the loxP sites. There are numerous Cre mice available expressing Cre in a tissue specific or inducible manner, so that the deletion can be controlled temporally and spatially (Kos, 2004; Sauer, 2002).

The targeting strategy for LP10/Hennes is presented in Figure 22. The full sequence of the targeting vector is available in the digital supplement. Cre mediated recombination results in the deletion of LP10 exons 2 and 3 and in the deletion of the Hennes exons 1 – 3 including the putative promoter region. It is anticipated that a functional knock-out of both genes is achieved, because LP10 ORFs are located in exon 3 and because Hennes expression is inactivated due to promoter disruption.

3.3 Phenotypic analysis of LP10/Hennes knock-out mice

Total knock-out mice were obtained from crossing mice carrying the conditional allele with an X-linked Cre-deleter mouse ubiquitously expressing Cre under the control of CMV promoter (Schwenk et al., 1995). The genomic deletion was verified by PCR. Heterozygous and homozygous LP10/Hennes mutant mice were viable and indistinguishable from their wild type littermates in survival rate, general appearance and behavior. Mating male knock-out mice with wild-type mice revealed that they are not sterile. However, to determine if the knockout gave more subtle effects, a series of further tests was performed.

Stained sections of testis and epididymis from knock-out mice and their wild-type littermates showed no difference between both groups (Figure 23). The histological analysis indicated that spermatogenesis proceeded normally and that all types of germ cells were produced. Mature spermatozoa were present in the epididymal ductus. Data on testis weight, epididymis weight, total sperm count per epididymis, total round cell

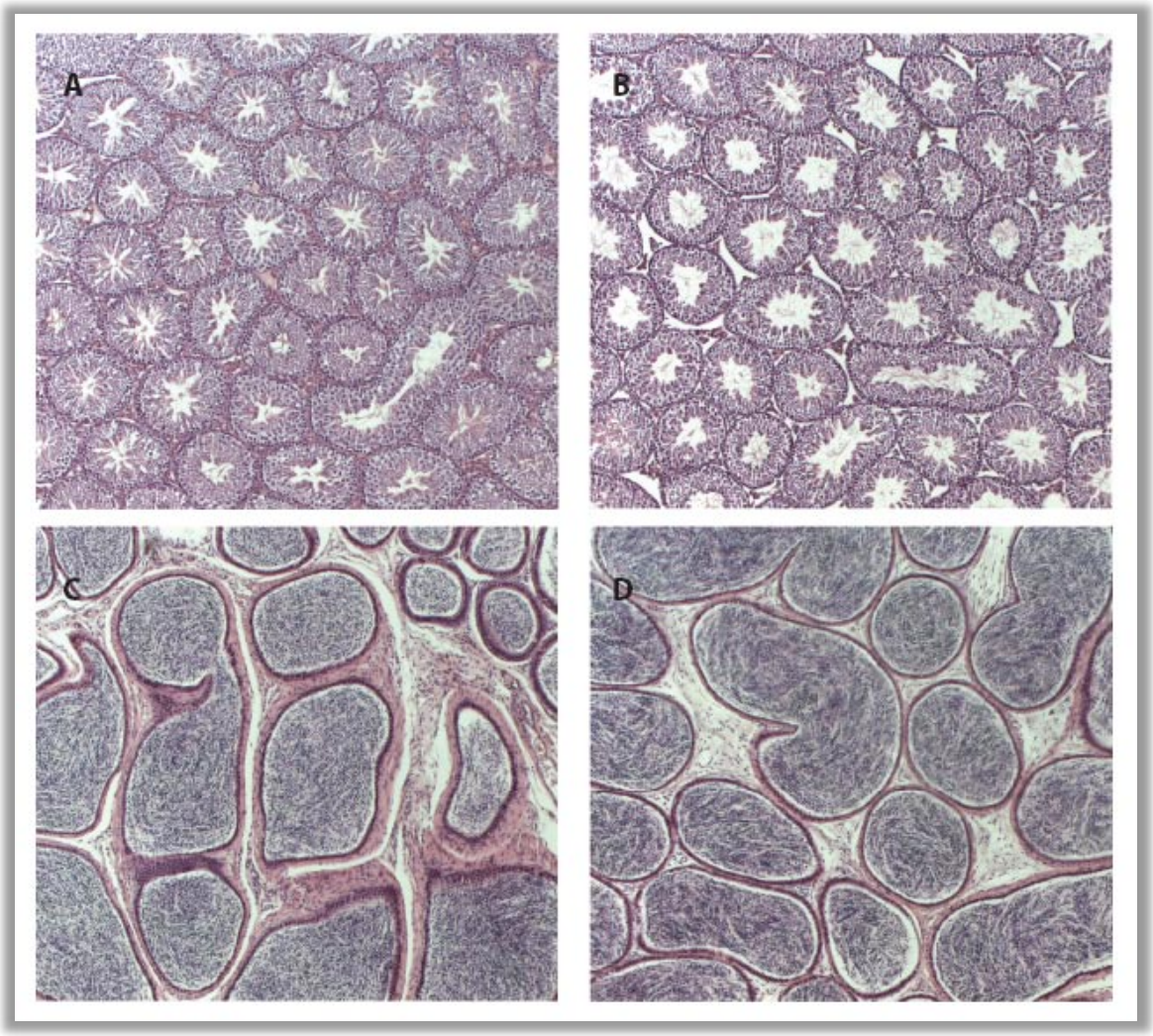


Figure 23 Histological analysis of wild type (A+C) and LP10/Hennes knock-out (B+D) mice. The upper two pictures show testis and the lower two show epididymis. Tissue sections were stained with hematoxylin and eosin and photos were taken at 50 fold magnification.

count per epididymis and sperm viability comparing 6 knock-out and wild-type mice were collected (Table 3). This revealed a tendency for the testes and epididymis of knock-out mice to have a higher mass, but the relatively large standard deviation points out that a larger sample size is required to verify this observation. Values of the total sperm varied in the single measurements due to the preparation method, but a difference between knock-out and wild-type mice was not observed. The total number of so called round cells has to be considered as well when looking at the total sperm cell count.

	wild-type	knock-out
testis weight [mg]	**79.8 ± 7.2**	**91.2 ± 9.4**
epididymis weight [mg]	**35.5 ± 13.6**	**26.5 ± 2.3**
total sperm count [x10^6]	**1.108 ± 0.941**	**1.156 ± 0.566**
total round cell count [x10^6]	**0.719 ± 0.409**	**0.506 ± 0.455**
sperm viability [%]	**63.3 ± 12.5**	**58.5 ± 3.4**

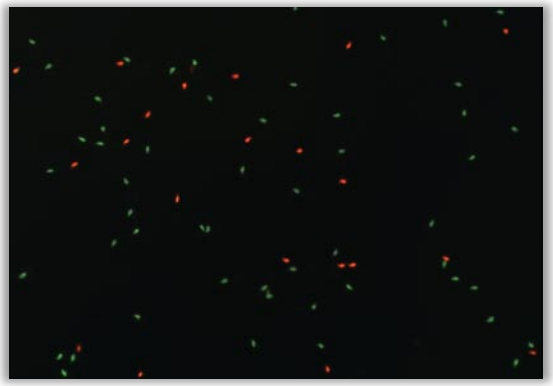


Figure 24 Example picture of fluorescently stained living spermatozoa. Alive cells are stained green and dead cells are stained red. Picture was taken at 400 fold magnification.

Table 3 Measurement of key data from the male reproductive system of wild-type and LP10/Hennes KO mice. Values were obtained by building mean and standard deviation of data from 6 mice.

Round cells defines all cells and particles different from spermatozoa that are found in the fluid analyzed (immature sperm cells, leukocytes, debris, etc.). They are assessed in order to control for the quality of the preparation and for healthiness of the animals. Their number is in an acceptable range although it also varies remarkably in the different animals indicated by large error bars. Sperm viability was measured by staining of living spermatozoa simultaneously with different fluorescent colors for dead and live cells (example picture in Figure 24). The sperm viability was not found to be different in the animals compared.

As a further test, the sperm motility was compared in wild-type and knock-out mice. Spermatozoa were categorized into four classes according to a protocol of the World Health Organization (World Health Organization, 1992):

a) fast progressing cells (≥20 µm/s at RT)
b) slow progressing cells (≤20 µm/s at RT)
c) non progressing, but moving cells (≤5 µm/s at RT)
d) immotile cells

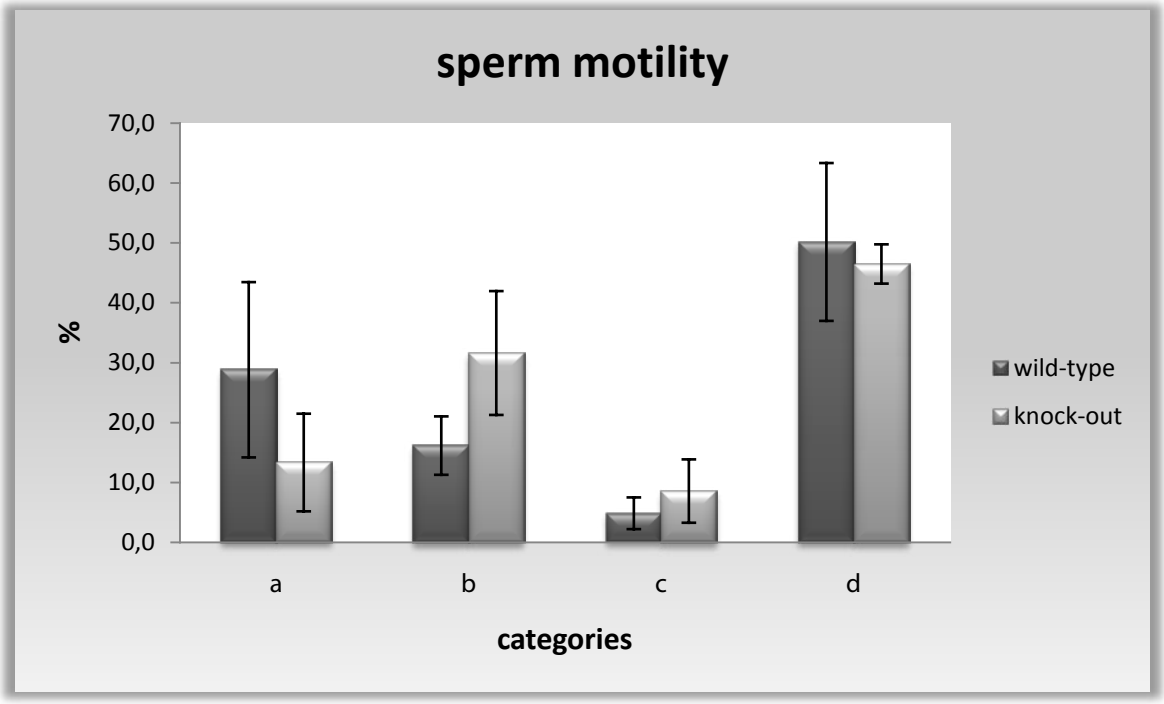


Figure 25 Sperm motility of wild-type and LP10/Hennes KO mice. Spermatozoa were counted by microscopy and categorized into 4 classes (a – d) according to their motility. The categories are described in the main text. At least 200 cells from different visual fields were counted per animal and the percentage of every category was calculated. The graphic displays mean and standard deviation from 5 wild-type and 6 knock-out mice.

The results of the experiment are presented in Figure 25. It is evident that sperm motility is shifted from class “a” towards class “b” in knock-out mice. The reduction in number of class “a” spermatozoa in the knock-out mice includes cells with morphological anomalies that can be seen with phase contrast microscopy in living cells (Figure 26A). They are characterized by a nicked tail which hinders normal cell movement. Such irregularities were also observed in some wild-type mice and indeed a few knock-out mice only mildly exhibited this phenotype. The large error bars reflect this inconsistency. Spermatozoa that were fixed on slides and stained do not show abnormalities in knock-outs compared with those from wild-type mice (Figure 26B+C).

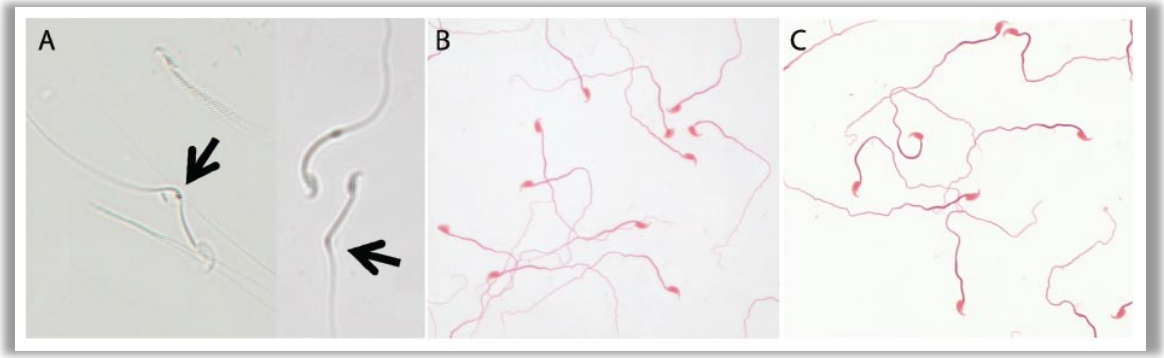


Figure 26 Sperm morphology. (A) Some living sperm cells of wild-type and LP10/Hennes KO mice showed morphological abnormalities. Sperm tails were nicked at the position of the annulus (arrows). Pictures in A were made using phase contrast microscopy. (B + C) Eosin staining of fixed sperm cells did not show any differences between wild-type (B) and KO (C) animals. All Pictures were taken at 400 fold magnification.

3.4 Expression analysis of LP10/Hennes knock-out mice

Microarray technology provides the possibility to track expression levels of large numbers of genes (Schena et al., 1995). The availability of LP10/Hennes double knock-out mice allows the analysis of the regulatory influence of both genes on expression of other genes in testis by comparing their global transcriptional status with wild-type controls. Affymetrix Genechip arrays are widely used and well established microarray tools for large scale expression profiling (Dalma-Weiszhausz et al., 2006; Heber and Sick, 2006). Genechip Mouse Genome 430 2.0 arrays cover the analysis of over 39000 transcripts on a single array. They were used to compare testicular gene expression of LP10/Hennes knock-out mice and wild-type mice. Testis RNA from 4 individual animals of each group was used for the analysis. The RNA was controlled for quality and integrity with Agilent Bioanalyzer Chips prior to the experiment (data is available in the digital supplement). Microarray data was normalized with the Mas5 method (Affymetrix, 2001; Gautier et al., 2004) and subsequently analyzed with the SAM-tool (significance analysis of microarrays). SAM uses non-parametric statistics to identify significant differences in gene expression (Tusher et al., 2001). For each gene an expected score is calculated based on the variance in expression values of all input samples. Another score for each gene is calculated based on the factual difference in

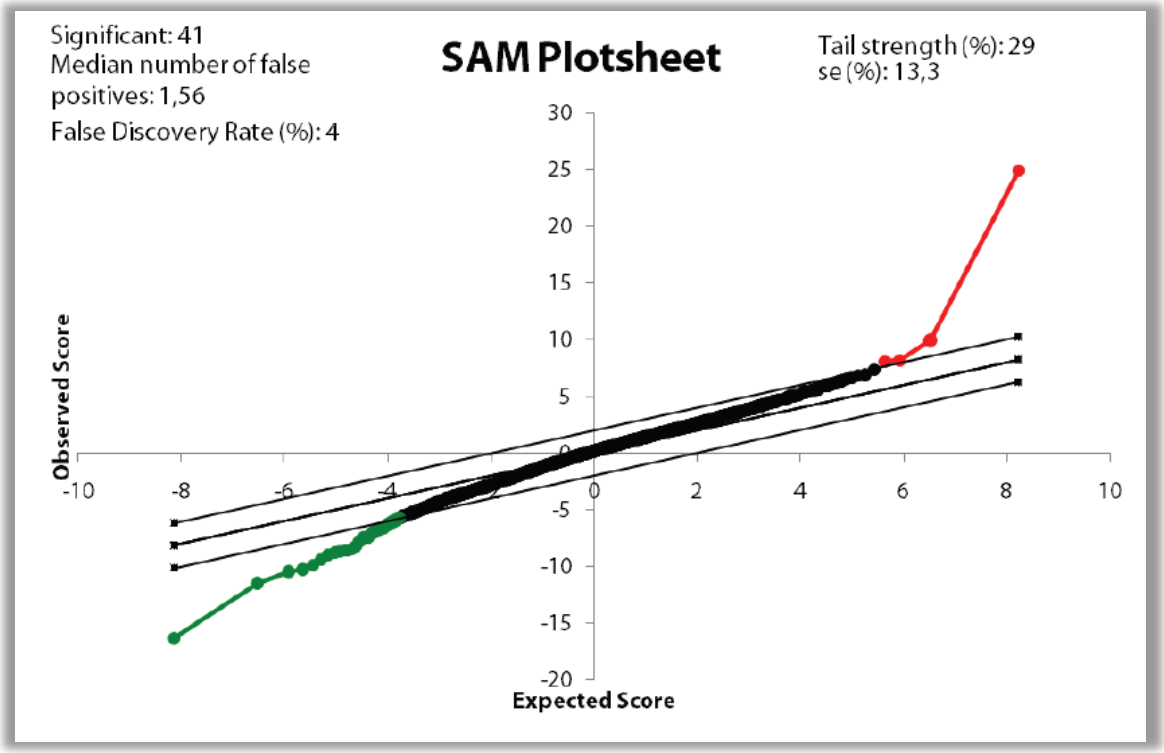


Figure 27 SAM-plot of LP10/Hennes KO testis microarray data. 4 knock-out animals were compared with 4 wild-type animals. 37 genes in knock-out mice (green) and 4 genes in wild-type mice (red) are significantly differentially expressed.

variance of both comparison groups. Both scores are plotted against each other and genes whose scored differ from each other are regarded as differentially expressed. An overall false discovery rate (FDR) is calculated that displays the expected percentage of falsely discovered genes. The more both scores diverge the lower is the expected FDR-rate. Figure 27 shows a SAM plot of the analyzed data. With a FDR rate of 4%, 4 genes are significantly higher expressed in wild-type mice while 37 genes are higher expressed in the knock-out mice. The gene names of the significant genes were retrieved from their probe IDs using the NetAffx tool (Liu et al., 2003). The table with significant genes is provided in Supplement B. The gene being most significantly higher in wild-type mice is LP10 as a result of its absence in the knock-out mice. Note that Hennes is not represented on the chip, because a probe set is missing. According to the data it can be assumed that either LP10 or Hennes or both of them might have a

repressing influence on expression of some genes in testis. However in most cases the change in expression was relatively small. A gene ontology annotation of the genes up-regulated in knock-out mice, using the MGI Gene Ontology Term Finder (http://www.informatics.jax.org/gotools/MGI_Term_Finder.html), did not detect any particular biological process as being significantly affected (Supplement C). However, 19 genes belong to the class "unannotated". An interesting candidate is 1700112E06Rik, because it is a testis specific gene and 15 fold up-regulated in knock-out mice. The Ssty2 gene is a highly expressed gene located on the Y-chromosome and could therefore be interesting for further studies, although its expression fold change is only slightly affected by LP10 knockout.

3.5 The evolutionary origin of LP10

Orthologous genes of LP10 and Hennes are not known in other species, although the genomic region is syntenic among mice, rats and humans. LP10 and Hennes did not evolve by duplication, because no paralogous genes can be found in mice. The genomic locus can be aligned with syntenic regions of rat and human. Northern blot analysis comparing testis RNA in different closely related genera of Murinae was performed to address the question of whether LP10 evolved de novo (Figure 28). Expression was detected with a LP10 probe amplified from Black6 in three subspecies of *Mus musculus* (*musculus*, *domesticus*, *castaneus*) and moreover in *Mus spretus* (LATASTE, 1883) (very weak expression), *Mus macedonicus* (PETROV & RUZIC, 1983) and *Mus cypriacus* (CUCCHI, 2006). It is missing in *Mus spicilegus* (PETÉNYI, 1882), *Mus famulus* (BONHOTE, 1898), *Mus caroli* (BONHOTE, 1902), *Apodemus sylvaticus* (LINNAEUS, 1758) and *Rattus norvegicus* (BERKENHOUT, 1769). Additionally, the down-regulation in *M. m. musculus* compared to *M. m. domesticus* that was originally observed in microarray screens is confirmed by the experiment. The blot was stripped and re-hybridized with a tubulin probe for loading control. For another control, house mouse and rat testis RNA was blotted and hybridized with a LP10 probe amplified from rat genomic DNA (Figure 29).

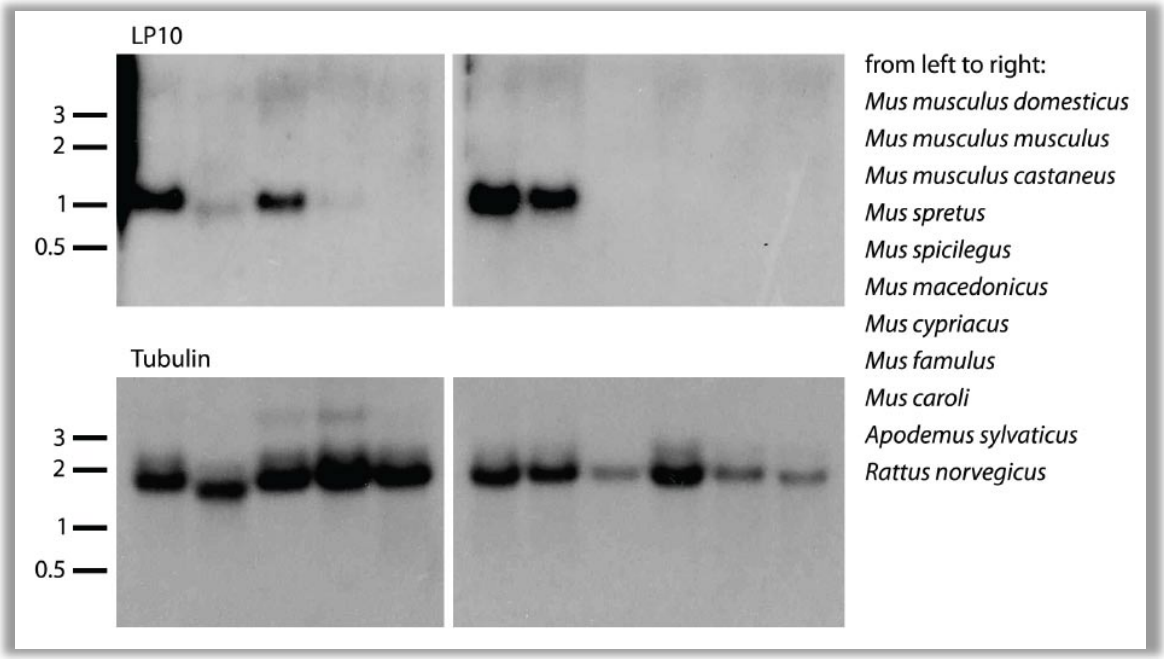


Figure 28 Northern blot analysis of testis RNA from different genera of Murinae. The blot was hybridized with a mouse LP10 probe and re-probed with a tubulin probe. Size marker bands represent kb.

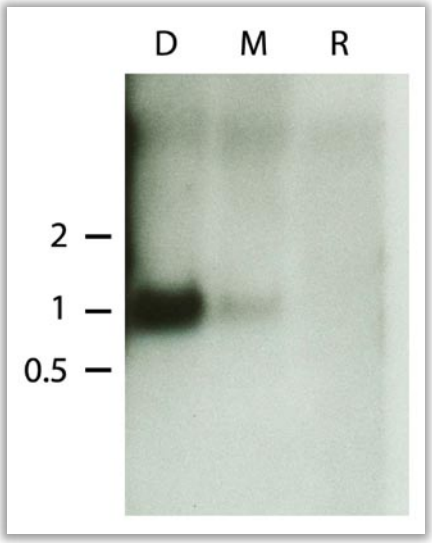


Figure 29 Northern blot with testis RNA from *M. m. domesticus* (D), *M. m. musculus* (M) and *Rattus norvegicus* (R) hybridized with a rat probe. Size marker bands display kb.

The result implies that the sequence homology is not too distant to detect related species. The rat probe binds to mouse RNA but not to rat RNA. The results from the species comparisons can be related to previously published phylogenetic trees shown in Figure 30 (Chevret et al., 2005; Guenet and Bonhomme, 2003). *Mus cypriacus* is a recently discovered species (Cucchi et al., 2006) and was not included in these phylogenetic analysis, but it is known that it diverged together with *Mus spicilegus* and *Mus macedonicus* (Macholan et al., 2007b). Based on the Northern blot results and the phylogenetic data it becomes evident that LP10 emerged de novo 2 million years ago.

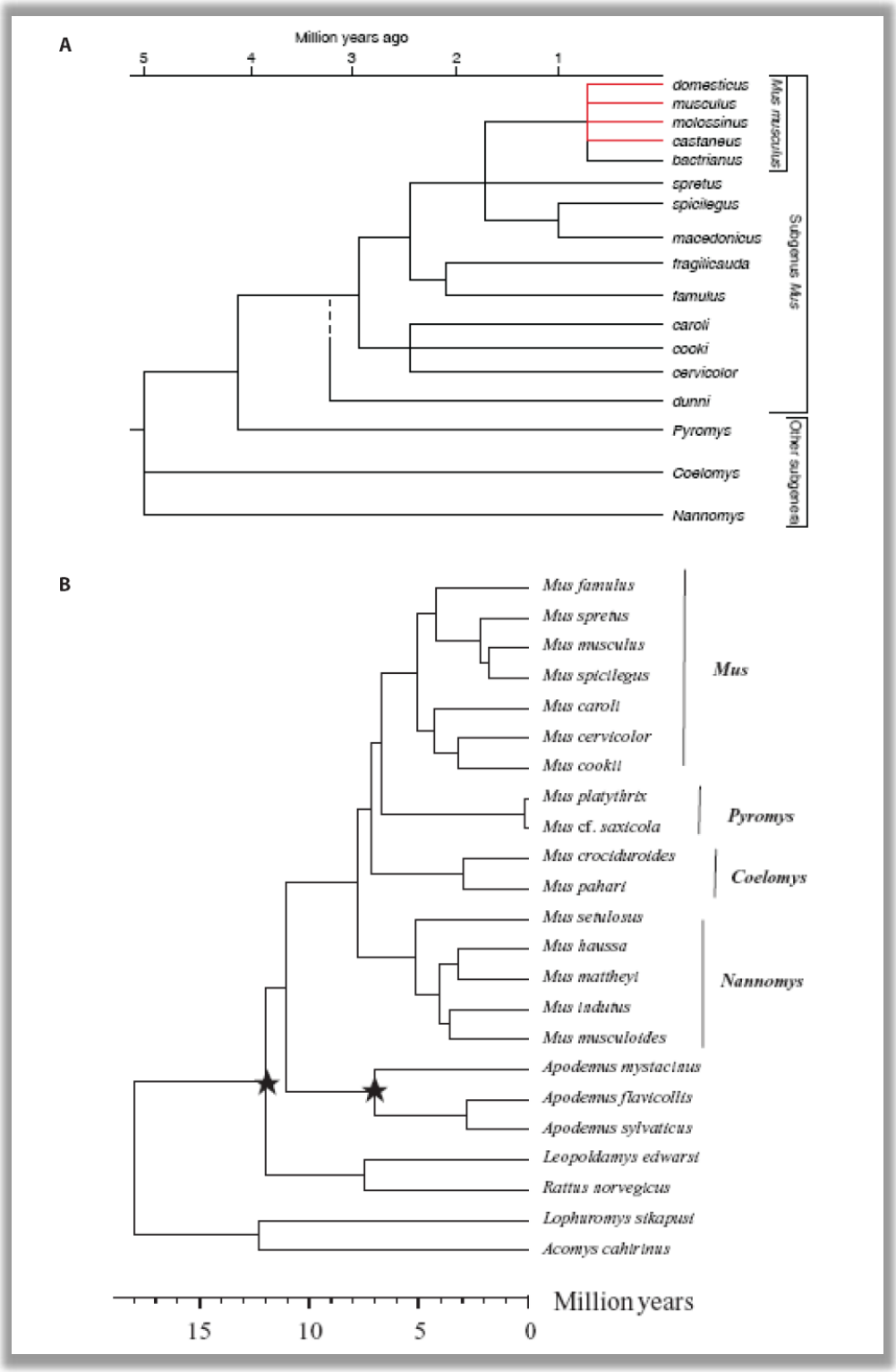


Figure 30 Mouse phylogenetic trees. Pictures are taken from Guenet and Bonhomme (2003) (A) and from Chevret et al. (2005) (B).

3.6 Summary

The LP10/Hennes locus is located in the centre of a selective sweep valley in the *Mus musculus musculus* lineage. Additionally, LP10 was previously found to be differentially expressed between *Mus musculus musculus* and *Mus musculus domesticus.* Therefore, the locus was chosen for a functional follow up study. LP10/Hennes double knock-out mice were generated. No significant difference between knock-out and wild type mice was found in a phenotypic analysis of the male reproductive system. A microarray analysis comparing gene expression of wild-type and knock-out mice revealed that LP10/Hennes might have a small influence on regulation of other genes. A general biological function of the differentially expressed genes was not found, but some of the genes are testis specific genes which might be interesting candidates for further analysis. A Northern blot analysis of different species of Murinae yielded that LP10 is a de novo evolved gene with an age of approximately 2 million years.

4 Discussion

Further investigation of two genes implicated in recent adaptation - Mkk7 and LP10/Hennes - was presented in this thesis. Experiments to understand their evolutionary changes were performed. Both genes had different premises in terms of previously existing knowledge about their function and therefore different emphasis was placed on the analyses. In the following the results of the genetic characterizations of LP10/Hennes and Mkk7 are discussed in the broader context of the evolution of genes, regulatory interactions and providing experimental evidence for natural selection.

4.1 LP10 and Hennes

4.1.1 De novo evolution of LP10

Phylogenetic expression comparisons of different wild mice showed that LP10 is possibly only 2 million years old and evolved de novo from non-coding DNA in a common ancestor of *Mus musculus*, *Mus spretus*, *Mus macedonicus* and *Mus cypriacus*. Widely accepted theories about the origin of new genes comprise gene duplication including retrotransposition, modifications mediated by transposon insertions, exon shuffling, horizontal gene transfer (mostly in Prokaryota, rare in Eukaryota) and gene fusion while de novo evolution is thought to be extremely rare (Long et al., 2003). According to the neutral theory of molecular evolution (Kimura, 1979) the mechanism of creating new genes by gene duplication has been considered the singular and most important one (Ohno, 1970; Ohta, 1988). The accordant “waiting model” suggests that gene duplication relaxes the constraint on one of the duplicated copies. In succession, the other copy can evolve neutrally before it acquires a new function. In this regard, de novo evolution of genes from non-coding DNA is an unexpected finding. Generally, the evolution of new genes is importantly influenced by positive selection (Long et al., 2003) as it was also observed for LP10.

Future analysis should address the question of whether Hennes also evolved de novo. A Hennes Northern blot experiment was so far not possible in part because the RNA is

apparently only very weakly expressed, which did not allow recovery of a cDNA clone suitable for blot hybridization. An interesting question is whether LP10 and Hennes share the same cis-regulatory elements and whether the same mutation caused de novo evolution of both transcripts.

Just very few cases of de novo gene evolution have been described to date. A pioneering study recently identified 5 de novo genes by comparing closely related species of the Drosophila complex (Levine et al., 2006). Subsequent publications affirmed the existence of this phenomenon in Drosophila (Begun et al., 2007; Zhou et al., 2008). De novo genes found in Drosophila are short spliced transcripts with short ORFs (the translation status remains unclear) that occur predominantly in testis (Levine et al., 2006; Long et al., 2003; Metta and Schlotterer, 2008). Interestingly, this applies to LP10 as well. LP10 is the first gene described as de novo, which may have evolved from non-coding DNA in mammals. The age of 2 million years is also very young for a mammalian gene. In comparison, it seems therefore also plausible that de novo genes evolved even in humans after the chimpanzee human split that took place 4-6 million years ago (Hobolth et al., 2007).

LP10 offers an exciting opportunity to study the mechanism of how new genes are born in mammals. The fact that de novo genes are predominantly expressed in testis leads to the conjecture that testis promoters are less complex and need different regulatory requirements as in other tissues. Testis-bias maybe also linked with altered chromatin structure in post meiotic cells (Caron et al., 2005). 99% of male germ cell-specific genes are expressed post meiosis (Schultz et al., 2003). The fact that testis expresses the highest level of alternative splicing factors should be considered in this context as well (Grosso et al., 2008). Investigating the molecular mechanisms associated with de novo genes and other uncharacterized transcripts in testis will be of great importance for molecular biology in general and evolutionary biology in particular.

4.1.2 Function of LP10 and Hennes

A LP10/Hennes knock-out mouse line was screened for phenotypic anomalies. No significant differences were found between knock-out mice and wild-type controls. A tendency was noticed for knock-out mice to develop testis and epididymis of higher weights than controls, but more animals have to be analyzed in order to ascertain this finding. A higher testis weight in knock-out mice would be interesting with regard to the reduced expression of LP10 in the sweep population, because testis weight can generally be seen as a fertility factor. Another experiment showed a tendency towards reduced sperm motility in knock-out mice. This phenotype was correlated with abnormal sperm morphology. It occurred to different extents in knock-out and also in wild-type mice leading to inconsistent data. The measurement of sperm motility is critical, because sperm cell quality is very sensitive to various factors such as temperature and pH. Moreover the sample quality decreases rapidly with time after preparation. The analysis was performed with technical limitations and should be repeated using a different experimental setup. Different improved systems to evaluate sperm motility by special counting chambers for light microscopy (Imade et al., 1993; Makler, 1978) or by full automation (Agarwal and Sharma, 2007; Slott et al., 1993) are available. They should be considered for future experiments. Moreover, the space for mouse breeding was limited. Therefore, the analysis should be repeated in future and a larger sample size should be surveyed.

A clear functional role for LP10 and Hennes was not inferred by phenotypic analysis of knock-out mice, but a regulatory effect was observed by microarray experiments. The fold changes of the differentially expressed genes are not very large in general and significant grouping of the genes to shared biological functions was not found. A relatively small number of differentially expressed genes was detected. Thus, it is very unlikely that LP10 or Hennes act as transcription factors in a common sense, but they might indirectly contribute to transcriptional regulation as minor modifiers. This could be realized by modifying proteins involved in transcription, by binding to RNA or DNA or by altering the chromatin structure at certain loci. Nevertheless, some of the

differentially expressed genes should be considered for follow up-analysis. The gene 1700112E06Rik showed the highest fold change (15.48), but it is expressed at a very low level. LOC100042492 and Ubiquitin Specific Peptidase 42 are more strongly expressed and showed a moderate but clear change. Ssty2, a testis specific gene located on the Y-chromosome, showed a low fold change (1.13), but it is highly expressed and the expression values are consistent in both groups compared. The function of all these genes is not known and qRT PCR analysis should be performed to verify the expression differences before any debate about their role as LP10/Hennes downstream targets.

In summary, although the function of LP10/Hennes remains unclear, different aspects that can direct future avenues of study were identified. For example, analysis of testis weight, sperm motility and regulatory effects. Moreover, it is still an important question to determine whether LP10 is translated. Proteins of both LP10 ORFs have been expressed in *E. coli* and were used for antibody generation. But the Western blots have not yet yielded clear results (data not shown).

The overall goal of the experiment was to understand the function of LP10/Hennes in order to point out a functional process as possible target for selection. A direct assumption concerning this matter cannot be made yet. However, a role of LP10/Hennes in fertility would be particularly interesting. LP10/Hennes knock-out mice are not sterile, but a fertility effect of the genes may be more subtle. Normal sperm cells maybe out-compete sperm cells of LP10/Hennes knock-out mice. Direct comparison of reproductive success of knock-out mice and wild-type mice, for example by mate choice experiments or in vitro fertilization, would therefore be interesting.

4.2 Mkk7

4.2.1 Regulation and function of Mkk7 in wild mice testis

The regulation of the Mkk7 locus is very complex. Different isoforms are expressed in a tissue specific manner. Mkk7α1 is a testis specific variant of Mkk7 that was identified in *Mus musculus domesticus* but is absent in *Mus musculus musculus*. It is responsible

for a highly elevated level of total Mkk7 RNA in *Mus musculus domesticus* testis that has been previously associated with a selective event. Reporter gene experiments identified a single mutation that is responsible for enhanced activity of the Mkk7 alpha basal promoter driving the Mkk7α1 transcript. The strong expression of Mkk7α1 is extended towards late spermatids where no Mkk7 expression is found in *Mus musculus musculus*. A future experiment using the already existing reporter constructs could potentially verify the effect in of the mutation in vivo (see section 4.3). Northern blot experiments comparing Mkk7 alpha expression in different mouse species could show whether this isoform is newly evolved in *Mus musculus domesticus* or was lost in *Mus musculus musculus*. Genomic sequencing of Mkk7 alpha promoter sequences from different mouse species could correlate the mutation with the expression status in those species.

Interestingly, the usage of the alternative promoter implies a non-canonical polyadenylation mechanism absent in other transcripts. RNA processing including capping, splicing and polyadenylation are co-transcriptional processes that are interconnected and influenced by the transcriptional machinery (Proudfoot and O'Sullivan, 2002; Proudfoot et al., 2002). Other examples already showed the co-usage of alternative 5' and 3' UTRs (Winter et al., 2007; Yang et al., 2003a). Those alternative UTRs are believed to play an important role particularly in spermatogenesis (Kleene, 2001; Yang et al., 2003a). It is unclear whether mutations that were found in the proximity of the polyadenylation site in *Mus musculus domesticus* contribute to the non-canonical splicing. A qRT-PCR analysis approach of hybrids between *Mus musculus musculus* and *Mus musculus domesticus* could help to understand whether cis-regulatory changes are involved in the alternative polyadenylation process in *Mus musculus domesticus*.

Another interesting finding was the existence of a 5'CACACA motif in the Mkk7 alpha promoter. It is the only insulator known from the literature that is involved in the maintenance of spermatid specific expression and it was shown that it is able to repress

Mkk7 alpha promoter expression in non-spermatid cells. It would be an interesting subject for further analysis. In silico analysis could help to measure the abundance of the motif among testis specific basal promoters and reporter gene expression could address the question of whether this insulator is also capable to repress expression of other promoters which naturally do not contain the sequence.

The only known function of Mkk7 is the activation of JNK, however Western blot analysis showed that the JNK pathway is not activated in testis. In general, MAP signaling cascades are usually activated by certain stimuli and are not constantly active. In contrast, spermatogenesis is a constant process. It is very unlikely that a unique highly expressed variant of a gene evolves in testis whose protein product is not used. Therefore another function for Mkk7 in testis is probable. The analysis of the protein levels of Mkk7 in wild mice testis has so far yielded only equivocal results. Comparing low amounts of testis protein resulted in a different outcome than comparing high amounts of protein on the Western blot. It can be excluded that the samples were mixed, because the result was reproducible. At present one can only speculate that either an unknown systematic mistake occurred or an unknown biochemical effect is involved. Other protein samples were not available for testing, but the experiment should be repeated with more samples in the future.

However, other experiments gave more clear results. The synthesis of a novel peptide, ^{sp}Mkk7, by translation of a different reading frame was assumed. An antibody against this peptide specifically stained a band in Western blot experiments which had the expected size and was constantly stronger in *Mus musculus domesticus*. The experiment lacks so far the unequivocal proof that the band indeed reflects ^{sp}Mkk7, although this is likely. The attempt to generate a positive control by in vitro transcription/translation of Mkk7α1 cDNA using rabbit reticulocyte extract failed and was therefore not included into the presented work. The finding that different peptides are potentially translated from the same RNA would deserve closer attention, because this would contradict a common paradigm.

In summary, it was not possible to determine the molecular function of Mkk7 in testis. Immunoprecipitation of the peptide stained with the anti ^{sp}Mkk7 antibody and subsequent Edman degradation or mass spectrometry can be performed to identify the peptide. The same procedure can be applied to proteins that were co-precipitated with an anti Mkk7 antibody after cross-linking of natively isolated testis protein in order to identify potential interacting proteins of Mkk7. Conditional knock-out mice of Mkk7 are in preparation. They can boost the effort to explain the function of Mkk7 in testis and they will also be a very useful negative control in Mkk7 protein analysis. Moreover, the construction of a knock-out that constitutively deletes the Mkk7 alpha promoter without altering the other isoforms is part of ongoing work. These mice will answer the question of whether the Mkk7α1 transcript provides a benefit in reproductive success by comparing them with wild-type control mice or with wild mice.

4.2.2 Mkk7 is an example for promoter evolution

The emergence of a new promoter in Mkk7 is the first example of a single nucleotide mutation in a cis-regulatory region that changes expression, and has been a possible target for positive selection, in mammals. The question of how genetic changes contribute to adaptive evolution of form and function is a controversially discussed subject. A common theory that has been developed over decades suggests a predominance of cis-regulatory changes in evolutionary divergence (Britten and Davidson, 1969; Carroll, 2008; Davidson, 2001; Rockman and Stern, 2008; Wray, 2007), however, others have proposed that, while cis-regulatory changes are important, the role of coding changes is underestimated (Hoekstra and Coyne, 2007; Lynch et al., 2008; Wagner and Lynch, 2008).

The main argument in this debate is the role of pleiotropic effects due to genetic changes. The proponents of cis-evolution argue that protein changes are more likely to cause deleterious pleiotropic effects, because they are affecting each cell type in which the respective protein is made. The coding-evolution proponents question the idea that pleiotropic effects of protein changes are more likely to be deleterious than pleiotropic

effects of temporal and spatial gene expression changes. However, cis-regulatory changes affect more than temporal and spatial gene expression changes. They can influence different levels of regulation. A recent study showed that 58% of mouse protein coding genes have two or more promoters. 92.9% of those genes are predicted to use distinct start codons (Carninci et al., 2006). The existence of alternative promoters is particularly important in testis. Different UTRs in spermatogenic cells are believed to alter translational behavior in order to prevent cells from deleterious defects caused by high levels of protein synthesis due to increased and altered expression patterns (Kleene, 2001).

The Mkk7 example demonstrated how the change of a single base presumably leads to the generation of an alternative promoter causing a new 5'UTR that influences RNA processing and protein translation. Thus, it is difficult to disentangle which effect is more important in this case, the new promoter or the new peptide that is generated as a consequence of it. Functional studies will be required to test these alternatives separately.

4.3 Measuring expression of naturally occurring promoter variants

The identification of mutations which determine the expression strength of naturally occurring promoter variants is an important subject for evolutionary research. Therefore, a renilla luciferase expression vector was constructed, which allows expression differences to be detected in vivo through site directed recombination. This technology creates embryonic stem cells with single copy insertions into a defined position of the CollagenA1 locus. The plasmid can be used in combination with the C2 ES cell line which can be used for the direct generation of fully ES cell derived animals by injection into tetraploid blastocysts. So, the typing and backcrossing of chimeric mice is omitted and time can be saved. Different reporter gene constructs can all be integrated at the same position in the genome and are always single copy. This is an advantage in comparison with conventional transgenic mice and BAC-transgenic mice which require the laborious characterization of founder animals. Another great

advantage is that promoter fragments can be amplified by PCR or retrieved from BACs and can easily be cloned into the Luciflip plasmid within a single ligation step. This saves the time consuming construction of individual targeting vectors for different genes as it would be necessary in common gene targeting experiments. A large amount of different constructs can be generated and tested in a short time. It was shown that the reporter plasmid is also suitable for transient transfection in cell culture experiments. As mentioned in section 2.5.3, the "flip-in" system into the CollagenA1 locus was previously used for successful expression of genes in an inducible manner. Although it still has to be proven that it can be applied to induce luciferase expression driven by natural promoter variants in vivo, this system could be a useful tool for future candidate gene characterization. A potential candidate gene for reporter gene analysis is β-Defensin6 (Defb6). A synopsis about this candidate gene is presented in the addendum following the discussion part.

4.4 An integrated approach to provide experimental evidence for natural selection

The work described in this disquisition contributes to the long term goal of providing experimental evidence for natural selection. The initial point of this approach is the identification of candidate genes that are differentially expressed between closely related species of wild mouse and show significant signs of recent selection at the DNA level. Modern genomic research technology facilitates such screens. Secondly, the candidate genes have to be carefully validated. The expression pattern has to be characterized in detail and other factors that could be a hidden cause for the selection event have to be considered. For example, it is expedient to exclude protein sequence mutations in the sweep population. Furthermore, it is helpful to test whether the expression difference is due to cis-effects rather than to trans-effects. Multiple comparisons with related species as they were conducted for LP10 can help to relate the expression difference with different genotypes and to trace the evolutionary origin of the phenotype.

The next step is to understand the molecular genetic cause for the expression difference in a functional context. Candidate genes that fulfill the requirements described above are not chosen according to their functional properties. Decisions for genes that are biased towards existing knowledge about function may ostensibly help to explain fitness benefits, but they may also obstruct an objective view. Moreover, choosing genes according to objective parameters underlines the significance of the whole approach in the long run. However, the functional consequences of the expression differences need to be unraveled. This can be demanding as it was evinced by the examples given in this is work.

It is facilitating the analysis if the expression difference of the candidate genes is restricted to certain tissues. Testis is a suitable organ to find such genes. There are many testis specific transcripts. Those genes are acting in unique processes which can only be found in testis. Comprehending the function of testis specific genes is anyway of great importance for the understanding of evolution. The drawback of investigating testis specific genes is the lack of appropriate cell culture models. Thus, reporter studies to identify the mutation causing the expression difference may have to be performed in vivo in some cases. This can be done as discussed above.

Once the cis-regulatory variant that causes the expression difference is identified it can be attempted to show that it causes an increased fitness as well. House mice offer the advantage that they can be genetically modified in a directed manner using homologous recombination in embryonic stem cells. Therefore, it is possible to generate almost identical clones that differ in just one base position. Mouse genomes can be manipulated in a way that a single mutation is inserted without any remaining traces of foreign DNA or other alterations (Cohen-Tannoudji and Babinet, 1998). Mouse lines that differ in just one position can be analyzed for fitness differences in direct comparisons. Releasing the mice into large enclosures can simulate natural conditions. So, it can be measured whether a mutation provides a selective benefit, for example an increased reproductive success. In some cases is would certainly be desirable to look in wild strain

backgrounds. Therefore, the establishment of wild-mouse derived embryonic stem cell lines is aspired.

This work has built a platform for further studies investigating the adaptation of Mkk7 and LP10 and the approaches and findings can be applied to other such candidates. The described methods offer a very promising approach to prove examples of natural selection as they were suggested by Charles Darwin (Darwin, 1859).

5 Addendum

Beta-Defensin6 - A potential candidate for future analysis

Mammalian β-Defensins comprise an important family of innate host defense peptides with pleiotropic activities (Patil et al., 2005). More than 30 different β-Defensins are encoded in the mouse genome. They evolved by gene duplication and have been the target of numerous selection events leading to a high degree of variation (Maxwell et al., 2003; Semple et al., 2006). Beside their role as antimicrobial peptides they have various functions that are not well understood. For example, the dog β-Defensin103 (orthologous to mouse Defb14) is involved in coat color determination (Candille et al., 2007). The sperm coating protein DEFB126 facilitates penetration of cervical mucus in Macaques (Tollner et al., 2008). A role of human β-Defensin2 in intestinal wound healing is presumed (Otte et al., 2008) and human β-Defensin3 is involved in activation of antigen-presenting cells (Funderburg et al., 2007). However, the expression properties of β-Defensins are as variable as their functional roles and their genomic copy numbers. Tissue specificity and regulation of gene expression differs drastically between different paralogs of β-Defensins (Kaiser and Diamond, 2000).

Defb6 was previously identified in a region that has been subject to a selective sweep in *Mus musculus domesticus* (Ihle et al., 2006). The amino acid sequence does not differ among the subspecies compared indicating that selection has possibly targeted the regulatory sequences of this gene. Defb6 is expressed in various epithelia, in throat and oral cavities, but predominantly in skeletal muscle tissue (Yamaguchi et al., 2001). Expression in skeletal muscle is unique among those β-Defensins for which expression has been analyzed. Over-expression of Defb6 in transgenic mice impairs lifespan and growth and results in a number of muscle abnormalities, as seen in human muscular dystrophy (Yamaguchi et al., 2007). Expression analysis using SYBRgreen based qRT-PCR comparing the Defb6 expression in esophagus of *M. m. domesticus* and *M. m. musculus* demonstrated that expression levels were somewhat significantly different

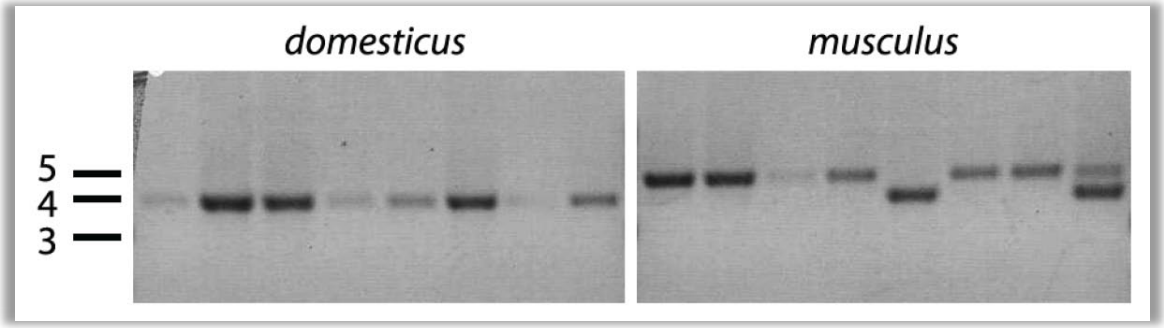


Figure 31 Sequence length polymorphism of Defb6 promoter. Size standard displays kb.

(Ihle, 2003). The expression measurement in esophagus was repeated with an ABI TaqMan-assay. This technology is more accurate than SYBRgreen PCRs. A significantly lower esophagus expression in *M. m. domesticus* was observed with a fold change of 29.3. In order to find cis-regulatory differences related to the expression difference, the promoter region of Defb6 in 8 individuals of *M. m. domesticus* and *M. m. musculus* was amplified by PCR for sequencing. Subsequent gel electrophoresis revealed a distinct length polymorphism (Figure 31). 6 *M. m. musculus* animals were homozygous for the insertion, one animal was heterozygous and another one didn't have the insertion at all. On the other hand all *M. m. domesticus* animals were homozygous for the short allele. Sequencing of the PCR fragments revealed an 829 bp insertion (-2527 base pairs relative to the transcription start) to be the cause of the length polymorphism (Supplement D). Repbase – a database of known repetitive elements - was searched using the CENSOR online tool to identify similar entries (Kohany et al., 2006). The Defb6 promoter insertion matched the sequence of mouse specific LTR retrotransposon RLTR13B2. Retrotransposon insertions have important effects on the regulation of gene expression (Buzdin et al., 2005; Hasler et al., 2007). Thus, the Defb6 promoter insertion would be an interesting subject for reporter gene analysis using the Luciflip system. A preliminary test to answer the question of whether the expression difference is also present in skeletal muscle would be possible since a cell culture model for this cell type is available (Yaffe and Saxel, 1977). If it could be shown that the transposon causes differential expression in muscle cells, it could be argued that purifying selection has eliminated the transposon allele in *M. m. domesticus*

due to the deleterious effects of high Defb6 expression in skeletal muscles that were described in a transgenic mouse model. On the other hand, if this is not the case, because the expression difference is indeed restricted to epithelial cell types in esophagus or elsewhere, it would be very difficult to explain why a lower expression in *M. m. domesticus* can be the target of positive selection. Moreover, sequencing of the Defb6 promoter region revealed the existence of several other sequence differences between *M. m. musculus* and *M. m. domesticus* (Supplement D). Investigating their potential contribution to the expression difference complicates the scenario. The existence of synergistic effects is likely and would handicap the identification of mutations that are responsible for the expression difference and how they act.

6 Materials and methods

6.1 Animals

6.1.1 Wild mice

The *Mus musculus musculus* animals used in Northern blotting, Western blotting and in situ hybridization belong to the JPC 2821 wild derived strain from Czech Republic. Wild-caught *Mus musculus musculus* animals from Austria were used for quantitative real time PCR, RACE and additionally for the total-Mkk7 Northern blot. All used *Mus musculus domesticus* animals were derived from wild-caught mice from the Cologne/Bonn area. *Mus musculus castaneus* mice belong to the CIM-strain. These animals were kindly provided and previously described by Ruth Rottscheidt (Rottscheidt, 2007). *Mus spretus* animals used for Northern blotting derived from laboratory breedings of wild animals caught in Spain (Voolstra et al., 2007). Wild mice were kept in cages in open shelves under non-sterile conditions.

6.1.2 Laboratory mice

C57BL/6J (Black6) mice, LP10 conditional and knock-out mice and Cre-deleter mice were sterilely kept in isolated ventilated racks in the mouse facility of the Institute for Genetics, Cologne.

6.2 Reagents

6.2.1 Chemicals and Kits

If not specified separately, standard chemicals were obtained from Applichem (Darmstadt), Sigma-Aldrich (St. Louis), VWR international (Darmstadt) or Carl Roth (Karlsruhe). MinElute PCR Purification Kit, QIAquick PCR Purification Kit (both Qiagen, Hilden) and NucleoSpin Extract II (Macherey-Nagel, Düren) were used for PCR purification and DNA gel-extraction. QIAprep Spin Miniprep Kit, QIAGEN Plasmid Midi Kit, QIAGEN Plasmid Maxi Kit (all from Qiagen, Hilden) and

NucleoSpin Plasmid QuickPure (Macherey-Nagel, Düren) were used for Plasmid purification. TOPO-TA Cloning Kit, Zero Blunt TOPO Cloning Kit (both from Invitrogen, Carlsbad) and pGEM-T easy Vector system (Promega, Mannheim) were used to clone PCR products into plasmids for amplification and preservation purposes. All other kits that were used are listed separately in the context of procedure descriptions. The ssRNA ladder (New England Biolabs, Ipswich) was loaded on RNA gels to measure band sizes. Protein mass was estimated with the SeeBlue Plus2 Pre-Stained Standard (Invitrogen, Carlsbad). Band sizes and masses in DNA gels were determined with one of the Low Range, Middel Range or HighRange FastRuler DNA Ladders (Fermentas, St.Leon-Rot) or with 2-Log DNA Ladder (New England Biolabs, Ipswich).

6.2.2 Enzymes

Restriction enzymes, arctic phosphatase and proteinase K were obtained from New England Biolabs (Ipswich). T4-DNA-Ligase was received from Roche (Basel). All other enzymes that were used are listed separately in the context of procedure descriptions.

6.2.3 Antibodies

rabbit-anti-Mkk7 (Cell Signaling, Danvers), working dilution 1:1000

rabbit-anti-SAPK/Jnk (Cell Signaling, Danvers), working dilution 1:1000

rabbit-anti-phospho SAPK/Jnk (Cell Signaling, Danvers), working dilution 1:1000

mouse-anti-α tubulin (Sigma-Aldrich, St. Louis), working dilution 1:4000

rabbit-anti-^{sp}Mkk7 (Eurogentec, Cologne), working dilution 1:1000

anti-rabbit-IgG, HRP-linked (Cell Signaling, Danvers), working dilution 1:10000

anti-mouse-IgG, HRP-linked (Cell Signaling, Danvers), working dilution 1:2000

anti-DIG, AP-linked (Roche, Basel), working dilution 1:3000

6.2.4 Primer list

All primers were obtained from Metabion (Martinsried). The following list contains all primers (P) that are named in the course of procedure descriptions.

Primer name	Primer sequence
P26	CTGCAAGCAAATGAACGTGC
P27	GGCCTGACATCGAGGAGGAT
M13forward	CGCCAGGGTTTTCCCAGTCACGAC
M13reverse	TCACACAGGAAACAGCTATGAC
P34	CCTGGGTTCTTTTCCAACG
P35	CTAGCAAAATAGGCTGTCCC
P36	GGATCCCTGCTCGTTCTTCAGCACGC
P37	CGCCTCCTGGATCACTACAAGTA
P45	GTGTTTGTCTACCCCGGACT
P47	CGTGGAAGACTTTGTCCAGA
P49	AATTAACCCTCACTAAAGGGGAGCATCGAGATTGACCAGA
P50	TAATACGACTCACTATAGGGGCTCGGATGTCATAGTCAGG
P55	CTCCTGACATACGGTCCTGT
P58	TGCCCATGTAGGTCACTAAG
P83	ACCTTCTCTCGCACACATGCT
P85	GTAATGTTTAGCTGCTTGTGGTG
P86	AGGCCAGAGTATTGGGATGC
P107	AGTCTGGAGGAGTCTCCCACTT
P113	GAATTCTTACAGATCCTCTTCAGAGATGAGTTTCTGCTCCTGCTCGTTCTTCAGCACGC
P114	CCAAAGTGTGAAGTGTTCTGCCT
P121	CCTGTAGGGAGGAATAAATCCTGG
P122	CACTGAGTTCCTGGAGGTGTTTAGT
P123	AGCACACTCCTTTGGAGCAA
P124	GTCCTGTGATCCATCCAATAGCT
P125	TCAGAACTGAGGGACTGAGGTG
P126	AATGAATAAATGAGCATGTATGAATCTG
P142	GGGTTGACCAAAGGTTTGAA
P144	TAATGACCAGAGAAAGGCAAGAG
P145	CAGTGCCTTATATAGTGCCAGGT
P149	TCATGTCCTTGCCATAGTTCCTA
P171	TTTAAAAAGGGAAGAGGGCAACA
P177	GTCTTCCGTTCACAGTGTCTGTC
P180	CTTTGTCCTAGGATGGCTTTGTG
P182	AGTCTGGAGGAGTCTCCCACTTC
P183	AGGGAAGGGTTTTATTCTCAAGC

P184	AAAGATCACAGGAAGAGACCAAAGT
P199	AGGCCCAAAGACAGTAGGCTCAG
P263	CTCATGGCAGCGGCCAGC
P270	AATTAACCCTCACTAAAGGGAACACAGAGGGCAGAAGTCC
P271	TAATACGACTCACTATAGGGGGGCTCTGCTAGGCTGCTGG
P288	ACGGTATCGATAAGCTTGCAG
P290	GAGGGTCCATGGTGATACAAG
P292	GATCTGTAGGGCGCAGTAGTCC
P293	GAAAGACCGCGAAGAGTTTGTC
P302	CTGGCCTGACTGACCTTCAC
P303	TAATACGACTCACTATAGGGCTCTGGGTATCTCCAGTTCCTG
P304	CTACCCCGGACTGACGGGTG
P305	TAATACGACTCACTATAGGGGTGGGCTGATATCCAAGTTGA
P306	TAGAACCCTATTCCCAACCCTGT
P312	CCCCCTCCATATAACATGAATTT
P316	CAATTGTTACAGATCCTCTTCAGAGATG
P317	GAATTCGCCACCATGGCTTCCAAGGT
P318	CTCGAGTGACCAACTACTTTTCACTATTGCTG
P319	CAAGCTGTGAAGGTCAGTCAGG
P320	TGGTGGACAAGCTGGATCTAGAAAGGAAGAGGAAGCACT
P321	CTCTTCCTTTCTAGATCCAGCTTGTCCACCATGACC
P322	CTCGAGCGCGTTACATAACTTACGGTAAA
P323	CTCGAGCAAAACAAACTCCCATTGACG
P324	CTTCACAGCTTGATCATCTTCC
P325	AAGAAACTCCCAGGAAGGAAA
P326	CTCCCTGGAGCAGAAGCTGT
P327	GTGGGCTGATATCCAAGTTGA
P328	CCCATCAACCTTGTTCACAC
P329	GTCAGGTACCCTGTCTGCTTC
P330	TGACTATGACATCCGAGCTGA
P331	CAAAGTCCGTCTTGCAGTTCT
P332	CACAGCTTCATCAAGCACTATG
P333	TGCTGACTCAGGACTCCACTA
P334	AGGACAGAAAGTAGGGGGTTC
P335	CAGAGTGGGAGGGATTAGGA
P338	AGTTGAGAAACCTTGCTGGAA
P342	CCCTCTGACCTCCTCCTCAG
P343	AGGGTTTTATTCTCAAGCGTCT
P344	TTGTCTTTGTCTACTGCCCTCT
P345	TGGTGACAGAGGGAAGGACTTG
P346	CAGTAGTCCAGGGTTTCCTT

P347	GAATTCCTCGATCCAGACATGATAAGATA
P348	GATGTACGGGCCAGATTTAC
P349	CTTTACCAACAGTACCGGATT
P352	TAGTCTCGTGCAGATGGACAGC
P353	CTCTTGCCGGACTTACCCATT
P354	TTAGAATGGGAAGATGTCCCTTG
P355	CTACACCCTGGTCATCATCCTG
P356	CCTGGCTCACAAATACCACTGA
P357	CGAGGTGCCGTAAAGCACTAA
P358	GCGTGCTTCTCGGAATCATAGT
P360	CATCTGCCCTTCTTCAGGTAGC
P361	CCTTCCACTGAGGGTAGAGTGG
LOH forward	TGCCAAATCACCCTGCTTGC
LOH reverse	TGTGCAAGCTGTAACCATCC
ROH forward	AGCCATAGCCTTGTCCAGAG
ROH reverse	CAGCTGCTTCTATTGGAAAGG
Cre forward	CGCATAACCAGTGAAACAGCAT
Cre reverse	GAAAGTCGAGTAGGCGTGTACG
rat LP10 forw	GTGAGGTCAGCCTGAGGAAG
rat LP10 rev	CCTGGACAGCAAGCCTAGAA
tubulin for	GTCTCCAGGGCTTCTTGGTTTT
tubulin rev	ACGCTTGGCATACATCAGATC

6.3 DNA samples

Wild mouse genomic DNA was used to obtain population sequence data and to amplify promoter fragments for reporter gene construction. DNA from *Mus musculus musculus* animals (Czech Republic) and *Mus musculus domesticus* animals (Cologne/Bonn area) was kindly provided by Meike Teschke and Sonja Ihle. All individuals were independently collected using a sampling strategy that eliminates the chance of collecting related animals by ensuring a minimal distance of 1 km between each collected mouse (Ihle et al., 2006). Rat genomic DNA was kindly provided by the department of Human Genetics (University of Cologne).

6.4 RNA Samples

Testis RNA from *Mus spicilegus*, *Mus macedonicus*, *Mus cypriacus*, *Mus famulus* and *Mus caroli* was kindly provided from Bettina Harr. Testis RNA from *Apodemus sylvaticus* was kindly provided from Fabian Staubach. Testis RNA from *Rattus norvegicus* was obtained from Applied Biosystems/Ambion (Austin). Other RNA samples were freshly isolated from mouse organs (see below).

6.5 Procedures

All laboratory techniques that are not specified explicitly were handled according to common laboratory procedures that were previously described together with the composition of standard buffers and solutions as they were used for this work (Sambrook et al., 1989).

6.5.1 Dissection of mice

For the isolation of mouse organs, mice were sacrificed by cervical dislocation. Then they were wetted with 70% ethanol for disinfection and to avoid that hair spreads onto the organs when the mice are cut open. A cut was made into the abdominal wall from which on the skin was cut open without damaging the muscle layer below. The muscle layer was then cut open with a clean pair of scissors and the organs were taken out. For the storage of the organs, they were shock frozen in liquid nitrogen and kept at –80°C.

6.5.2 Extracting RNA from Mouse tissue

Mouse organs were homogenized in Trizol Reagent (Invitrogen, Carlsbad) using the TissueRuptor system (Qiagen, Hilden). RNA extraction was performed by following the Trizol Reagent manufacturer's manual. RNA pellets were dissolved in pure nuclease free water (Applied Biosystems/Ambion, Austin). Photometric measurement of RNA concentration was performed with NanoDrop 1000 (Fisher Scientific, Hampton). Samples were stored at -80°C.

6.5.3 Extracting protein from mouse tissue

Mouse organs were homogenized in RIPA buffer (0.2 M NaCl, 1% TritonX-100, 0.1% sodium deoxycholate, 0.1% SDS, 50 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 tablet of complete protease inhibitor cocktail – EDTA free (Roche, Basel) per 50 ml) using the TissueRuptor system (Qiagen, Hilden) at 4°C. The lysate was separated from fat and debris by centrifugation at 4°C. Protein concentration in the cleared lysate was measured according to Bradford (Bradford, 1976). Samples were mixed with denaturing sample buffer (Laemmli, 1970) and stored at -20°C.

6.5.4 Extracting genomic DNA from Mouse tissue

Mouse tissue was lysed in 500 µl homogenization buffer (80 mM EDTA, 100 mM Tris-HCl pH 8.0, 0.5% SDS) with 10 µl proteinase K (10mg/ml) by shaking at 56°C over night. The lysate was cleared by centrifugation and genomic DNA in the supernatant was precipitated with the salt and alcohol method. DNA was whashed and resuspended in 10 mM Tris-HCl pH 8.0 and stored at 4°C.

6.5.5 Polymerase chain reaction

Two different polymerases were used for polymerase chain reactions (Mullis et al., 1986): Qiagen Multiplex PCR Kit (Qiagen, Hilden) and Phusion Hot Start High-Fidelity DNA Polymerase (Finnzymes, Espoo). The two polymerases are abbreviated with PCR Kit and Phusion in the following. For PCR Kit reactions 10 ng of template DNA were utilized in a 10 µl total reaction volume containing 5 µl of PCR Kit Master Mix and 0.2 µl of 10 µM dilutions from each primer. Cycling protocol: 15 min initial activation at 95°C, 40 cycles of 30 s denaturation at 94°C, 90 s annealing and 1 min / kb elongation at 72°C followed by 10 min final extension at 72°C. Phusion PCR reactions were composed of 4 µl 5x Phusion HF Buffer, 10 mM dNTPs, 0.5 µM per primer, 0.6 µl DMSO, 0.2 µl Phusion Hot Start DNA Polymerase and 200 ng of template DNA in a total volume of 20 µl. Cycling protocol: 30 s initial denaturation at 98°C, 30 cycles of 10 s denaturation, 30 s annealing and 30 s / kb elongation at 72°C followed by 10 min

final extension at 72°C. A standard annealing temperature of 63°C was used for most reactions. Annealing temperature was optimized by gradient PCR for some of the primer pairs. Thermal reactions were run in PTC-225 Peltier Thermal Cyclers (MJ Reasearch, Waltham) or in Gene Amp PCR System 9700 (Applied Biosystems, Foster City).

6.5.6 DNA sequencing

DNA sequencing reactions were composed of 0.5 µl sequencing primer (10 µM), x ng DNA, 1.5 µl 5x Big Dye sequencing buffer and 1 µl BigDye terminator mix version 3.1 (both from Applied Biosystems, Foster City) and brought to a final volume of 10 µl with pure water. For direct sequencing of plasmid DNA, 200 ng of purified DNA was used as template. Otherwise, DNA fragments to be sequenced were pre-amplified by PCR. Subsequently, 0.12 µl exonuclease I (20000 U / ml, New England Biolabs, Ipswich), 0.45 µl shrimp alkaline phosphatase (1 U / µl, Promega, Mannheim) and 2.43 µl of pure water were added to 10 µl of a PCR reaction. The mix was incubated for 20 min at 37°C and heat activated by 70°C for 15 min in order to dephosphorylate any dNTPs and remove remaining primers. 10 ng of thusly “EXO-SAP” treated DNA was used as template in sequencing reactions. Thermal cycling was run in Gene Amp PCR System 9700 (Applied Biosystems, Foster City) at following conditions: 1 min at 96°C followed by 30 cycles of 10 s at 96°C and 255 s at 60°C. Sequencing reactions were analyzed with an ABI 3700 DNA capillary sequencer. Sequence data was further analyzed using the CodonCode Aligner program (CodonCode Corporation, Dedham) and the Vector NTI software (Invitrogen, Carlsbad).

6.5.7 Sequencing of wild mouse genomic regions

Mkk7 intron1 and flanking regions from 8 individual animals of *M. m. musculus* and *M. m. domesticus* and from one Black6 animal were PCR-amplified by the PCR Kit using the primer pair P45 / P306. The resulting fragments were sequenced with primers P45, P47, P58, P55 and P306.

Mkk7 non-canonical polyadenylation sites and flanking regions from 12 individual animals of *M. m. musculus* and *M. m. domesticus* were PCR-amplified by the PCR Kit using the primer pair P360 / P361. The resulting fragments were sequenced with the primers P360 and P361.

Beta-Defensin6 promoter regions from 8 individual animals of *M. m. musculus* and *M. m. domesticus* were PCR amplified by Phusion and with the primer pair P144 / P145. The resulting fragments were sequenced with the primers P144, P121, P122, P123, P142, P124, P125, P126, P145. The retroposon in *M. m. musculus* animals was additionally sequenced with primer P149.

Sequence data was edited using the CodonCode Aligner program (CodonCode Corporation, Dedham) by annotating homozygous and heterozygous SNPs and indels and removing low quality sequence stretches. The Bioedit sequence editor was used to generate alignments (Hall, 1999).

6.5.8 RACE

Determining the 5’ and 3’ ends of transcripts was achieved via the RACE technique (rapid amplification of cDNA ends). RNA was reverse transcribed into cDNA, while linkers were added exclusively to those RNA molecules that had an intact 7-methylguanosine-cap (5’ RACE) or an intact polyA tail (3’ RACE) (Frohman et al., 1988; Volloch et al., 1994). The RACE cDNA was used as template in a subsequent PCR reaction using one primer that binds to the linker and another gene specific primer. RACE procedure was conducted using the GeneRacer Kit (Invitrogen, Carlsbad) according to the manufacturer’s manual. P50 was used as gene specific primer in 5’ RACE. P49 was used as gene specific primer in 3’ RACE. RACE-PCR reactions were performed with the PCR Kit and pooled as described in section 2.3.1. Pools were cloned in PCR cloning vectors and transformed into chemical competent NEB5alpha *E. coli* cells (New England Biolabs, Ipswich) which were subsequently streaked out on growth plates. Single clones were picked (5’ RACE: 48 clones per subspecies; 3’ RACE: 60

clones per subspecies) into 100µl water and colony PCR using the PCR Kit was run with M13 forward and M13 reverse primers. The PCR fragments were sequenced with both M13 primers and some 3'RACE clones had to be re-sequenced with P184.

P83 served as gene specific primer to amplify Mkk7 variant 3 (section 2.3.1, page 16, Figure 6) in gradient RACE-PCR using the PCR Kit. All reactions of the PCR gradient were pooled and the RACE-PCR fragments were retrieved from a gel after electrophoresis. PCR fragments were directly sequences with the primers P180, P85, P86, P338, P107 and P171.

All sequence data was analyzed with the CodonCode Aligner program (CodonCode Corporation, Dedham) and different isoforms were counted. The sequence raw data is attached in the digital supplement.

6.5.9 Generation of cDNA

Generation of cDNA was achieved by reverse transcription of 2 µg total RNA using 200 ng of random hexamers (Fermentas, St.Leon-Rot) and Superscript III Reverse Transcriptase (Invitrogen, Carlsbad) during a 60 min incubation at 50°C according to the manufacturer's instructions. cDNA samples which were compared in subsequent qRT-PCR experiments were reverse transcribed in parallel and equally treated reactions starting from RNA samples that were carefully controlled for equal concentrations.

6.5.10 Quantitative real time PCR with SYBR green

An Austrian *M. m musculus* sample and a German *M. m. domesticus* sample were compared using primer pairs that bind at different positions to Mkk7 cDNA: P326 / P327 to exon 1, P324 / 325 to exon2, P328 / P329 to exons 4 and 5, P330 / P331 to exons 10 and 11, P332 / P333 to the 5' region of exon 13 (upstream of variant 2 splice donor), P334 / P335 to exon 13 (in between the variant 2 splice donor and alternative polyadenylation site of Mkk7α1), P344 / 345 to 3' region of exon 13 (spanning variant 2 splice acceptor) and P342 / P343 to exon 14. Each amplicon is ~100 bp in length and

was amplified in parallel triplicates for each animal. SYBR-Green reaction was set up with the QuantiFast SYBR Green PCR Kit (Qiagen, Hilden) according to the manufacturer's manual and run with an ABI 7900 HT cycler (Applied Biosystems, Foster City). Cycle threshold (Ct) values were obtained from the ABI PRISM software (Applied Biosystems, Foster City). The mean of the Ct values from the triplicates were built and relative expression was calculated using the formula 2^{-Ct}.

6.5.11 TaqMan Assay for Defb6 quantification

TaqMan Gene Expression Assays (Applied Biosystems, Foster City) were used to quantify Defb6 expression in esophagus of *M. m. musculus* and *M. m. domesticus*. cDNA of 2 Czech *M. m. musculus* animals and of 3 German *M. m. domesticus* animals were used as templates in TaqMan qRT-PCR reactions. All samples were analyzed with an assay for Defb6 (Assay number Mm00651498_m1) and an assay for TBP as endogenous control (Assay number Mm01277042_m1). Each qRT-PCR reaction contained 6 µl TaqMan Gene Expression Master Mix (Applied Biosystems, Foster City), 0.6 µl TaqMan Gene Expression Assay and 5.4 µl cDNA (diluted 1:10). Every reaction was run as triplicate in parallel with an ABI 7900 HT cycler (Applied Biosystems, Foster City). Ct values were obtained from the ABI PRISM software (Applied Biosystems, Foster City). The Ct mean from the triplicates were built. For every individual animal the Ct of the endogenous control TBP was subtracted from the Ct of Defb6 resulting in delta Ct (ΔCt). The mean of ΔCt was formed for both subspecies and the relative fold change (Fc) was calculated according to the formula:

$$Fc = 2^{-(\Delta Ct_{domesticus} - \Delta Ct_{musculus})}$$

6.5.12 Histological techniques

Freshly prepared testes were incubated over night in 4% paraformaldehyde in PBS. The organs were paraffinized by consecutive incubations in different solutions: (at 4°C) 2x 1 h Formaldehyde, 1.5 h 70% EtOH, 1.5 h 80% EtOH, 1.5 h 96% EtOH, 3x 1h EtOH, 2x 1.5 h xylene and (at 60°C) 2x 2h Paraffin. Paraffinization was processed by the TP

1020 Tissue processor (Leica, Wetzlar). Organs were embedded in Paraffin using the EG 1150H device and cut into 7 µm sections with a RM225 microtome (both from Leica, Wetzlar). Sections were floated in a water-bath to spread the tissue (Leica, Wetzlar), transferred onto superfrost plus slides (Menzel-Gläser, Braunschweig) and dried on a warm plate (Leica, Wetzlar).

Sections were deparaffinized by incubation of 2x 10 min in xylene for subsequent hematoxylin-eosin (HE) staining. Afterwards they were hydrated in an alcohol serial (100%, 95%, 90%, and 70% EtOH for 5 minutes each). The blue staining of the nuclei was achieved by incubation for 3 minutes in hematoxylin followed by blueing under running tap water. The connective tissue is stained red by incubation in Eosin for 3 – 5 minutes. Afterwards, the slides are dehydrated in series of increasing alcohol content (70%, 90%, 95% and 100% EtOH for 5 min each) and terminated with clearing for 10 min in xylene. Sections were mounted with Canada balsam (Merck, Darmstadt).

6.5.13 In situ hybridization

In situ detection of Mkk7 RNA was performed by hybridization with a digoxigenin (DIG) labeled probe (Tautz and Pfeifle, 1989). For probe generation, a fragment spanning Mkk7 exons 5-10 was amplified from testis Black6 cDNA with primers P49 and P50 using the PCR Kit and cloned into a PCR cloning vector. The correctness of the insert was controlled by sequencing with M13 primers. The DNA fragment was re-amplified with primers P49 and P50 from a pure plasmid clone. Reverse transcription to generate a DIG labeled probe was set up by adding 200 ng of purified PCR product to 2 µl DIG RNA Labeling Mix, 2 µl transcription buffer, 2 µl T7 polymerase, 0.5 µl RNase inhibitor (all 5 from Roche, Basel). Pure water was added to the reaction mix to obtain a final volume of 20 µl. The reaction mix was incubated for 2 h at 37°C followed by a treatment with 1µl Turbo DNAse for 15 min at 37°C to remove the DNA template. The probe was precipitated with the salt and alcohol method, washed and re-suspended in 40 µl of 50% formamide diluted in nuclease free water (Applied Biosystems / Ambion, Austin).

All buffers and tools that were used for the following procedure were kept RNAse free. Paraffinized sections (see 6.5.12) were dewaxed in xylene for 2x 10 min, washed for 5 min in ethanol, rehydrated in a series of decreasing ethanol concentration (95%, 90%, 70%, 30%; 3 min each) and washed for 5 min in PBS before postfixing them for 1 h in 4% paraformaldehyde in PBS. After postfixation the tissue was washed in PBS for 2x 5 min and partially digested with 10 µg / ml proteinase K in 100 mM Tris-HCl pH 7.5 for 10 min at 37°C. Digestion was stopped with 0.2% glycine in PBS. 2x 5 min washing in PBS was followed by 15 min incubation in 0.1 N HCl and another 2x 5 min washing in PBS was performed previous to blocking of positively charged amino acids by 0.25% acetic anhydride in 0.1 M triethanolamine pH 8.0 for 10 min. Afterwards slides were washed for 5 min in PBS and for 5 min in pure water before pre-hybridization for 2 h at 65°C in pre-hybridization buffer (50% formamide, 5x SSC, 1x Denhardt's, 0.1% Tween-20). 1 µl of DIG labeled probe was diluted in 100 µl pre-hybridization buffer containing 400 ng tRNA (Sigma-Aldrich, St. Louis) and denatured at 70°C for 5 min. The hybridization mix was applied to the sections and covered with coverslips. Slides were incubated over night at 65°C in a moist chamber. Next day, the sections were washed in 50% formamide containing 5x SSC and 1% SDS at 70°C for 30 min and subsequently with 50% formamide containing 2x SSC and 0.2% SDS for another 30 min at 65°C. Afterwards the sections were washed for 3x 5 min in MABS (100 mM maleic acid, 150 mM NaCl, 0.1% Tween-20 and 2 mM levamisole; adjusted to pH 7.5 with NaOH). Samples were blocked with 1% blocking reagent (Roche, Basel) in MABS. Anti-DIG-AP antibody was applied in 1% blocking reagent in MABS by overnight incubation at 4°C. Next day, the sections were first washed 3x 10 min and then 3x 30 min in MABS. Subsequently, pH was adjusted by incubating for 3x 10 min in NTMLT buffer (100 mM Tris-HCl pH 9.5, 50 mM $MgCl_2$, 100 mM NaCl, 100 mM levamisole, 0.1% Tween-20). BM purple solution (Roche, Basel) was applied as substrate for the anti-DIG antibody coupled alkaline phosphatase. Tissue was stained until the desired degree of signal was observed. Slides were washed 1 min in water and mounted with Kaiser's glycerol gelatin.

6.5.14 Microscopy

Light microscopy including fluorescent and phase contrast microscopy was performed with an Axiophot 2 / Axioplan 2 microscope system (Carl-Zeiss, Oberkochen). Pictures were captured with an Axiocam camara and processed with Axiovision software (Carl-Zeiss, Oberkochen).

6.5.15 Northern blotting

Detection of RNA in Northern blotting (Alwine et al., 1977) was performed with radioactively labeled probes. Generation and cloning of a probe for total Mkk7 detection is described in section 6.12.13. The same fragment was used for total Mkk7 detection in Northern blot. Different additional probes were made for Mkk7 Northern blot experiments and were amplified from Black6 genomic DNA using the PCR Kit and the following primer pairs: Mkk7 exon1 with P304 / P305; Mkk7 exon2 with P302 / P303; 5' region of Mkk7 exon 13 specifically for Mkk7 variants 1 and 3 with P177 and P263; Mkk7 exon 14 with P182 / P183. The mouse LP10 probe was amplified on Black6 testis cDNA with PCR Kit and primers P270 / P271. LP10 rat probe was amplified on rat genomic DNA with PCR Kit and primers rat LP10 forw and rat LP10 rev. Alpha tubulin probe was amplified on Black6 testis cDNA with PCR Kit and primers tubulin for and tubulin rev. All probe fragments were cloned into PCR cloning vectors. The correctness of the inserts was controlled by sequencing with M13 primers. The probe fragments were re-amplified from a pure plasmid clone with the same primers that were initially used for probe amplification and extracted after gel electrophoresis.

Probes were labled with α^{32}P-dCTPs (Hartmann Analytic, Braunschweig) by the use of the Rediprime II DNA Labeling Kit (GE Healthcare Life Science, Little Chalfont) according to the manufacturer's manual. Labeled probes were cleaned up with MicoSpin S-200 HR columns (GE Healthcare Life Science, Little Chalfont) according to the manufacturer's manual.

10 µg of total RNA per sample were diluted in 15 µl nuclease free pure water (Applied Biosystems/Ambion, Austin) and mixed with 10 µl sample buffer (50% formamide, 5.18% formaldehyde, 2.5x MOPS, 0.1 mg/ml ethidiumbromide and 2.5x blue marker). Samples were heat-denatured for 5 min at 70°C and separated on an agarose gel (1.2% agarose, 0.666% formaldehyde, 1x MOPS). The RNA lanes were blotted through conventional upward blot onto a Amersham Hybond N+ membrane (GE Healthcare Life Science, Little Chalfont) by neutral transfer (20x SSC) over night. Membranes were baked for 2 h at 80°C and pre-hybridized in ExpressHyb (Clontech, Mountain View) at 65°C for 1h. Radioactively labeled probe was added to the pre-hybridized blot and hybridization took place over night at 65°C in a rotating oven. Next day, the blots were washed 10 – 40 min in 2x SSC containing 0.05% SDS at RT and subsequently washed for 5 – 30 min with 0.1x SSC containing 0.1% SDS at 50°C. After washing the blots were dipped in 2x SSC, sealed in a plastic bag and analyzed via autoradiography using Kodak Biomax-MS films (Kodak, Rochester).

6.5.16 Southern blotting

Detection of Mkk7 DNA after Southern blotting (Southern, 1975) was performed with a non-radioactively labeled probe. The Mkk7 probe is described in section 6.5.13. DNA was separated on a 0.8% agarose gel. The gel was equilibrated in 0.4 M NaOH and subsequently in transfer buffer (0.4 M NaOH, 0.6 M NaCl) before blotting through conventional upward blot onto a Amersham Hybond N+ membrane (GE Healthcare Life Science, Little Chalfont) by alkaline transfer over night. Membranes were baked for 2 h at 80°C and pre-hybridized in 10 ml ExpressHyb (Clontech, Mountain View) at 55°C for 1h. 1 µl DIG labeled probe was added into the solution and hybridization took place over night at 55°C in a rotating oven. Next day, the blot was washed 30 min in 2x SSC containing 0.05% SDS at RT and subsequently washed for 20 min with 0.1x SSC containing 0.1% SDS at 50°C. Membrane was dipped into 2x SSC. The reagents that are named in the following are described in section 6.5.13. The membrane was washed for 3x 5 min in MABS buffer and blocked for 30 min in 1% blocking reagent in MABS. Anti-DIG-AP antibody was applied in 1% blocking reagent in MABS for 1 h at RT.

Afterwards the blot was washed 3x 10 min in MABS without levamisole and equilibrated in NTMT. Bands were visualized with BM purple.

6.5.17 Western blotting

The extracts from 293 cells and 293 cells irradiated with UV for JNK and phospo JNK Westerns were obtained from New England Biolabs (Ipswich). 15µl of each extract was used for the experiment. The other protein extracts used are described in section 6.5.3. Proteins in 1x sample buffer were denatured for 5 min at 95°C and separated in discontinuous Tris/Glycine buffered SDS-PAGE (15%) prior to Western blotting. Protein gels were blotted onto Amersham Hybond-P PVDF membranes (GE Healthcare Life Science, Little Chalfont) using a Trans-Blot semi-dry transfer cell (Biorad, Hercules). After blotting, the membrane was blocked in 5% non-fat dried milk powder dissolved in PBS-T for 1 h at RT. Membranes were washed in PBS-T. The antibodies were applied in Can Get Signal Immunoreaction Enhancer Solutions (TOYOBO, Osaka). Primary antibody was applied over night in solution 1 at 4°C. Next day the membranes were washed 6x 5 min in PBS-T and secondary antibody was applied in solution 2 for 1 h at RT. Subsequently, membranes were washed 6x 5 min at RT. Chemiluminescence was catalyzed by horseradish peroxidase linked to the secondary antibody after incubation of the membranes with ECL detection reagent (GE Healthcare Life Science, Little Chalfont). The signals were visualized by exposing the membranes to Amersham Hyperfilm ECL (GE Healthcare Life Science, Little Chalfont).

6.5.18 Molecular cloning and testing of hrLuc-MYC

MYC-tagged hrLuc (Luc-MYC) coding sequence was obtained by PCR including the MYC-tag into the reverse primer sequence. Luc-MYC was amplified using the Phusion polymerase and pGL7.40 hrLuc plasmid (Promega, Mannheim) as template with primers P35 and P113 and cloned into a PCR cloning vector. The correct sequence was verified by sequencing with M13 primers as well as primers P37 and P26. The GFP coding sequence downstream of a CMV promoter in the phrGFPII-1 plasmid (Agilent Technologies, Santa Clara) was replaced with Luc-MYC coding region by molecular

cloning using the restriction sites EcoRI and HindIII. A control plasmid expressing the non-tagged hrLuc under the control of CMV was generated by retrieving the hrLuc coding sequence with KpnI and NaeI from the pGL7.40 hrLuc original plasmid. GFP was removed from phrGFPII-1 with KpnI and EcoRV. Luc-MYC-insert and CMV-vector were ligated.

C2C12 cells were grown in DMEM medium containing sodium pyruvate, non essential amino acids, L-glutamine, penicillin/streptomycin (all from Invitrogen, Carlsbad) and 10% fetal calf serum (PAN, Aidenbach) in 37°C incubation maintaining 5% CO_2. One day before transfection, 2.5 x 10^4 C2C12 cells were seeded into each well of 24-well plate and grown over night to ~80% confluence. 450 µg plasmid DNA and 2 µl Lipofectamine (Invitrogen, Carlsbad) were each mixed with 20 µl Opti-MEM medium (Invitrogen, Carlsbad) and incubated for 5 min at RT. In the meantime, the cells were washed with PBS and covered with 400 µl serum free medium. After incubation at RT for 5 min, both Opti-MEM preparations were mixed and further incubated at RT. After 45 min the mix was added to the cells and cells were placed in the 37°C incubator for the next 5 hours. After that period, 440 µl of medium containing 20% serum was added into each well. According to this procedure 3 wells were transfected with hrLuc control, 6 wells with Luc-MYC and 3 wells were treated with all reagents except DNA as blank control. Next day, each well was trypsinized (Invitrogen, Carlsbad) separately and the cells were transferred into the wells of a 96-well plate in 25 µl medium. Viviren (Promega, Mannheim) Luciferase substrate was prepared according to the manufacturer's manual and 2 µl of the solution were added to 98 µl medium. 5 µl of the diluted solution were added to each well and light emission was measured with a Luminoskan Ascent Luminometer (Fisher Scientific, Hampton). Mean and standard deviation were calculated for the technical replicates.

6.5.19 Molecular cloning and testing of Luciflip plasmid

The plasmid pBS31' (Beard et al., 2006) is available through the open biosystems project (http://www.openbiosystems.com/) and was obtained from a German distributor

(BioCat, Heidelberg). It contains a PGK promoter followed by an ATG and a FRT site which can compliment a promoter- and ATG-less resistance cassette in the ColA1 locus of C2 ES cells. It further contains a tet-inducible promoter (Schonig and Bujard, 2003) and a single EcoRI site downstream of that promoter. Therefore, cDNAs can be ligated into the vector by a single cloning step, transferred into the mouse genome and expressed in vivo after doxycycline treatment. An intron followed by a polyA signal enhances translational expression.

It was attempted to convert pBS31' into a promoter-testing system, by removing the tet-promoter and instead inserting the Luc-MYC coding sequence by maintaining the single EcoRI site. An upstream double polyA signal cassette protects the expression constructs from upstream transcriptional activity. The tet-promoter could not be removed without destroying this cassette because appropriate restriction sites are lacking. Therefore the double polyA signal cassette was amplified and a new EcoRI site was added to the 3' end using PCR Kit with the primer pair P346 and P347. The product was cloned into a PCR cloning vector and sequenced with M13 primer to verify the product. A fragment containing the double polyA cassette and the tet-promoter was removed from pBS31' and just the double polyA cassette was reinserted without the tet-promoter by molecular cloning of the double polyA PCR clone using the EcoRI site and an unique upstream ClaI site. The Luc-MYC cassette was re-amplified from a previously validated clone (see 6.5.18) using Phusion polymerase and the primers P316 and P317 to add an EcoRI site to the 5' end and an MfeI site to the 3' end. The fragment was cloned into a PCR cloning vector and sequenced with M13 primers, P36 and P37. No mutations were found. An EcoRI / MfeI fragment of a pure clone was ligated into the EcoRI site of the modified pBS31' plasmid which lacks the tet-promoter. EcoRI sites and MfeI sites are compatible but cannot be re-cleaved after ligation. Thus, just one EcoRI site upstream of Luc-MYC (downstream of the double polyA cassette) remained. Luc-MYC Kozak sequence was untouched by the procedure. The resulting plasmid was named Luciflip and contains the following elements in the given order: PGK-promoter, ATG, FRT, splice acceptor, double polyA signal, EcoRI site, Kozak, Luc-MYC, intron, polyA

signal. It was sequenced with the following primers: M13 forward, P26, P27, P34, P37, P288, P290, P292, P293, P312, P113, P314, P352, P353, P354, P355, P356, P357, P358. No mutations were found. The full sequence of Luciflip is presented in the digital supplement.

In order to test the function of Luciflip, a CMV promoter containing EcoRI sites on both ends was amplified from phrGFPII-1 plasmid (Agilent Technologies, Santa Clara) using PCR Kit and the primer pair P348 / P349. The PCR product was cloned into a PCR cloning vector and sequencing with M13 primers assured that no mutation occurred. The CMV promoter was retrieved from a pure plasmid and cloned into Luciflip by cutting and ligating the EcoRI site. Correct insert orientation was controlled by NcoI digest of the resulting plasmid and sequencing with P293 and P358.

NIH/3T3 fibroblast cells (Todaro and Green, 1963) were grown in DMEM medium containing sodium pyruvate, non essential amino acids, L-glutamine, penicillin/streptomycin (all from Invitrogen, Carlsbad) and 10% fetal calf serum (PAN, Aidenbach) in 37°C incubation maintaining 5% CO_2. One day before transfection, 3 x 10^3 NIH/3T3 cells were seeded with 70 µl medium into each well of 96-well plate and grown over night. Cells were co-transfected with Luciflip constructs and a pGL3 plasmid (Promega, Mannheim) which contains firefly luciferase under the control of SV40 promoter. Firefly luciferase was later used to normalize transfection efficiency. For transfection, 0.18 µl Fugene 6 reagent (Roche, Basel) was added to 4.82 µl serum free medium and incubated for 5 min at RT. Subsequently, 30 ng of Luciflip and 30 ng of pGL3 DNA were mixed and added to the medium containing Fugene 6. The mixture was incubated for 20 min and added to one well of the 96 well plate containing NIH/3T3. The transfection was performed for the Luciflip-CMV construct and for an empty Luciflip plasmid as blank control. Each transfection was performed in 8 replicates in parallel. Cells were incubated over night. Next day, firefly and renilla luciferase substrates were applied using the Dual-Glo Luciferase Assay System (Promega, Mannheim) according to the manufacturer's manual. Relative light units

were measured with a Mithras LB 940 Luminometer (Berthold Technologies, Bad Wildbad). For every well, the renilla luciferase signal was divided by the firefly luciferase signal to normalize transfection efficiency. For every construct, median and standard deviation was calculated from the 8 individual replicates.

6.5.20 Mkk7 alpha reporter assay

Fragments reaching from -487 to +43 relative to the transcription start of the Mkk7 alpha promoter were amplified from genomic DNA of *M. m. musculus* and *M. m. domesticus* using PCR Kit and the primer pair P318 / P319. The resulting fragments were cloned into a PCR cloning vector and sequenced with M13 primers. The only difference that was found in the plasmid clone was the naturally occurring SNP at position -84.

A two step PCR strategy using PCR Kit was pursued to generate fragments with a deleted insulator motive. Two separate PCRs with the primer pairs P318 / P320 and P321 / P319 were run on top of the cloned promoter fragment. The primers P319 and P320 bind just right upstream and downstream of the insulator sequence and are tailed with a sequence stretch which is homologous to the sequence on the opposite part exactly beyond the insulator. The other primer is one of the primers that were used in the first PCR. Thus, the promoter fragment is divided into two fragments each defined by an inner and an outer primer. The inner edges overlapped, but were lacking the insulator. Both PCR products were cleaned up and included into another PCR without primers. After 5 cycles, the outer primers P318 and P319 were added to the reaction and PCR continued as usual. The resulting product was cloned into a PCR cloning vector. Sequencing with M13 primers did not detect any unintended mutations. This was done for *M. m. musculus* and *M. m. domesticus* respectively.

Another version of all 4 fragments, which has an additional upstream CMV enhancer, was created. Therefore, the CMV enhancer was amplified with the primers P322 and P323 using the phrGFPII-1 plasmid (Agilent Technologies, Santa Clara) as template.

The resulting product was cloned into a PCR cloning vector and validated by M13 primer sequencing. Both CMV primers and the upstream primer that was used for promoter fragment generation (P318) are tailed with an XhoI restriction site overhang. CMV enhancer fragments were retrieved by XhoI digestion and ligated into the XhoI site of all 4 Mkk7 promoter fragments. The orientation of the insertion was controlled by EcoRI / NcoI double digest. The assembled fragments were cut out by EcoRI digestion and cloned into the EcoRI site of Luciflip. The orientation of the inserts was controlled with an XhoI digest. All 8 Luci-flip constructs were sequenced with the primers P293, P358 and P199 to control the inserts. No mutations were found.

The 8 expression constructs were transfected into NIH/3T3 cells and analyzed in 8 replicates each as described section 6.5.19.

6.5.21 Genotyping of LP10 knock-out mice

A 5 mm distal part of the tail was cut from each individual mouse for DNA isolation (see section 6.5.4). Genotyping of genetically modified mice was done by PCR using the PCR Kit. The primer pair LOH forward / LOH reverse flanks the loxP / FRT sites located downstream of LP10 in targeted animals. They generate a 351 bp fragment from wild type alleles, a 470 bp fragment from conditional alleles and no fragments from knock-out alleles. The primers ROH forward / ROH reverse are flanking the loxP / F3 sites located in LP10 intron1 in targeted animals. They generate a 326 bp fragment from wild type alleles, a 518 bp fragment from conditional alleles and no fragments from knock-out alleles. The primer combination LOH forward / ROH reverse generates a 436 bp fragment from knock-out alleles but no fragments from other alleles. The presence of Cre was tested with the primer pair Cre forward / Cre reverse which generates a 600 bp fragment from the Cre allele. Samples were genotyped by all kinds of those primer combinations run in parallel PCRs together with positive and negative controls (samples of known genotypes).

6.5.22 Phenotypic analysis of the male reproductive system

Male reproductive systems of 6 wild-type mice and 6 knock-out mice were phenotyped. The workflow was conducted according to a modified version of a previously published procedure (Behrens et al., 2002). Testis and epididymis were dissected from the right side of each mouse and weight was measured. The cauda epididymis was excised, immediately transferred in 250 µl human tubular fluid medium (Millipore, Billerica), punctured and placed at 37°C. In the meantime the weight of the right testis was measured again and subtracted from the total weight of epididymis and testis. The left testis and epididymis were excised, transferred into 4% PFA in PBS and incubated at 4°C over night for later histological analysis (see 6.5.12).

After 5 min of incubating the punctured right epididymis at 37°C, a 6 µl drop of medium with dispersed spermatozoa was transferred onto a warmed glass slide and covered with a 20 x 20 mm coverslip. Progressive motility was estimated by phase contrast microscopy at a magnification of 200x according to WHO (World Health Organization, 1992). At least 200 sperm cells per animal (but up to 700) in different areas of the slide were counted and the percentage of different classes of motility was calculated per animal. The overall classification of motility was expressed as the mean of the individual percentages from wild-type mice and knock-out mice.

Two more aliquots of medium with dispersed spermatozoa were taken for sperm morphology 5 min after puncturing the epididymis:

Sperm morphology was studied according to a protocol presented by the Jackson Laboratory: “A slide is prepared by spreading 5-10 ul of a sperm suspension across the surface. After drying, slides are fixed in 5% acetic acid in 95% ethanol for 3 min, air dried, stained in 5% Eosin Y (Sigma) for 4 min, rinsed in 70% ethanol, air dried, and mounted with 30 ml of Permount (Fisher). Slides are scored using bright field optics with a 60X objective. Abnormal sperm are those that exhibit aberration of the head (e.g. length-width ratio or shape), defective attachment of the head to the tail, defects of

tail (length, morphology) or persistence of the cytoplasmic droplet." (http://reproductivegenomics.jax.org/maleprotocol.html)

Sperm viability was determined 5 min after puncturing the epididymis by dual staining SYBR-14 in combination with propidium iodide (Garner and Johnson, 1995) using the LIVE/DEAD® Sperm Viability Kit (Invitrogen, Carlsbad) according to the manufacturer's manual. Cells were examined using fluorescence microscopy. At least 200 sperm cells were counted per animal and the percentage of green fluorescing spermatozoa was defined as viability.

The concentration of spermatozoa and round cells was measured 30 min after puncturing the epididymis, using a Neubauer hemocytometer. At least 16 small squares of the counting grid were taken into account.

6.5.23 Microarray analysis

Total RNA from 4 LP10 knock-out mice and 4 wild-type mice was controlled for quality by running an Agilent RNA 6000 Nano Kit with Agilent Bioanalyzer (Agilent Technologies, Santa Clara) following the manufacturer's manual. All samples showed RNA integrity numbers of beyond 8 and were used for microarray analysis. Agilent data is attached in the digital supplement. Preparation of biotinylated cRNA samples, hybridization of samples with Genechip Mouse Genome 430 2.0 arrays (Affymetrix, Santa Clara), washing of arrays using the fluidic station 450, staining with Streptavidin-Phycoerythrin and reading of data with an Affymetrix GCS 300 G scanner was conducted at the Cologne Center for Genomics. Microarray raw-data is attached in the digital supplement. Probe data from Affymetrix CEL-files was normalized with the MAS5 method (Affymetrix, 2001; Gautier et al., 2004) using the R-based bioconductor software (http://www.bioconductor.org/). The normalized probe set data was searched for differentially expressed genes with the significance analysis of microarrays (SAM) (Tusher et al., 2001). The resulting list of gene identifier was converted into

GeneSymbols using the NetAffx tool (Liu et al., 2003) and annotated according to gene ontologies with the MGI Gene Ontology Term Finder: (http://www.informatics.jax.org/gotools/MGI_Term_Finder.html).

7 Outside services and contribution of other people

- LP10 conditional knock-out mice were generated at TaconicArtemis GmbH (Cologne).
- Microarray experiments (Sample preparation, hybridization and scanning) were performed by the Cologne Center for Genomics. Most of this work was done by Christian Becker, Berlin.
- Daniela Häming took care of the laboratory mouse breedings, helped with their analysis, established Western blotting in the lab and participated in the LP10 Northern blot experiments under my supervision.

8 References

Abhyankar, M.M., Urekar, C., and Reddi, P.P. (2007). A novel CpG-free vertebrate insulator silences the testis-specific SP-10 gene in somatic tissues: role for TDP-43 in insulator function. J Biol Chem *282*, 36143-36154.

Abzhanov, A., Kuo, W.P., Hartmann, C., Grant, B.R., Grant, P.R., and Tabin, C.J. (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. Nature *442*, 563-567.

Acharya, K.K., Govind, C.K., Shore, A.N., Stoler, M.H., and Reddi, P.P. (2006). cis-requirement for the maintenance of round spermatid-specific transcription. Dev Biol *295*, 781-790.

Affymetrix (2001). Affymetrix Microarray Suite User Guide, version 5 edition edn (Santa Clara, CA, Affymetrix,).

Agarwal, A., and Sharma, R.K. (2007). Automation is the key to standardized semen analysis using the automated SQA-V sperm quality analyzer. Fertil Steril *87*, 156-162.

Alwine, J.C., Kemp, D.J., and Stark, G.R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc Natl Acad Sci U S A *74*, 5350-5354.

Araki, K., Araki, M., and Yamamura, K. (1997). Targeted integration of DNA using mutant lox sites in embryonic stem cells. Nucleic Acids Res *25*, 868-872.

Beard, C., Hochedlinger, K., Plath, K., Wutz, A., and Jaenisch, R. (2006). Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genesis *44*, 23-28.

Begun, D.J., Lindfors, H.A., Kern, A.D., and Jones, C.D. (2007). Evidence for de novo evolution of testis-expressed genes in the Drosophila yakuba/Drosophila erecta clade. Genetics *176*, 1131-1137.

Behrens, A., Genoud, N., Naumann, H., Rulicke, T., Janett, F., Heppner, F.L., Ledermann, B., and Aguzzi, A. (2002). Absence of the prion protein homologue Doppel causes male sterility. EMBO J *21*, 3652-3658.

Bhaumik, S., and Gambhir, S.S. (2002). Optical imaging of Renilla luciferase reporter gene expression in living mice. Proc Natl Acad Sci U S A *99*, 377-382.

Birgersdotter, A., Sandberg, R., and Ernberg, I. (2005). Gene expression perturbation in vitro--a growing case for three-dimensional (3D) culture systems. Semin Cancer Biol *15*, 405-412.

Blaise, R., Guillaudeux, T., Tavernier, G., Daegelen, D., Evrard, B., Mairal, A., Holm, C., Jegou, B., and Langin, D. (2001). Testis hormone-sensitive lipase expression in spermatids is governed by a short promoter in transgenic mice. J Biol Chem *276*, 5109-5115.

Bonhomme, F., Guenet, J.L., Dod, B., Moriwaki, K., and Bulfield, G. (1987). The polyphyletic origin of laboratory inbred mice and their rate of evolution. Biological Journal of the Linnean Society *30*, 51-58.

Boursot, P., Auffray, J.C., Britton-Davidian, J., and Bonhomme, F. (1993). The Evolution of House Mice. Annual Review of Ecology and Systematics *24*, 119-152.

Boursot, P., Din, W., Anand, R., Darviche, D., Dod, B., Von Deimling, F., Talwar, G.P., and Bonhomme, F. (1996). Origin and radiation of the house mouse: mitochondrial DNA phylogeny. Journal of Evolutionary Biology *9*, 391 - 415.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem *72*, 248-254.

Britten, R.J., and Davidson, E.H. (1969). Gene regulation for higher cells: a theory. Science *165*, 349-357.

Britton-Davidian, J., Fel-Clair, F., Lopez, J., Alibert, P., and Boursot, P. (2005). Postzygotic isolation between the two European subspecies of the house mouse: estimates from fertility patterns in wild and laboratory-bred hybrids. Biological Journal of the Linnean Society *84*, 379 - 393.

Buzdin, A., Vinogradova, T., Lebedev, Y., and Sverdlov, E. (2005). Genome-wide experimental identification and functional analysis of human specific retroelements. Cytogenet Genome Res *110*, 468-474.

Candille, S.I., Kaelin, C.B., Cattanach, B.M., Yu, B., Thompson, D.A., Nix, M.A., Kerns, J.A., Schmutz, S.M., Millhauser, G.L., and Barsh, G.S. (2007). A -defensin mutation causes black coat color in domestic dogs. Science *318*, 1418-1423.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., *et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet *38*, 626-635.

Caron, C., Govin, J., Rousseaux, S., and Khochbin, S. (2005). How to pack the genome for a safe trip. Prog Mol Subcell Biol *38*, 65-89.

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell *134*, 25-36.

Chang, L., Jones, Y., Ellisman, M.H., Goldstein, L.S., and Karin, M. (2003). JNK1 is required for maintenance of neuronal microtubules and controls phosphorylation of microtubule-associated proteins. Dev Cell *4*, 521-533.

Chang, L., and Karin, M. (2001). Mammalian MAP kinase signalling cascades. Nature *410*, 37-40.

Chevret, P., Veyrunes, F., and Britton-Davidian, J. (2005). Molecular phylogeny of the genus *Mus* (Rodentia: Murinae) based on mitochondrial and nuclear data. Biological Journal of the Linnean Society *84*, 417-427.

Cohen-Tannoudji, M., and Babinet, C. (1998). Beyond 'knock-out' mice: new perspectives for the programmed modification of the mammalian genome. Mol Hum Reprod *4*, 929-938.

Cooke, H.J., and Saunders, P.T. (2002). Mouse models of male infertility. Nat Rev Genet *3*, 790-801.

Cotta-de-Almeida, V., Schonhoff, S., Shibata, T., Leiter, A., and Snapper, S.B. (2003). A new method for rapidly generating gene-targeting vectors by engineering BACs through homologous recombination in bacteria. Genome Res *13*, 2190-2194.

Cucchi, T., Orth, A., Auffray, J.C., Renaud, S., Fabre, L., Catalan, J., Hadjisterkotis, E., Bonhomme, F., and Vigne, J.D. (2006). A new endemic species of the subgenus Mus (Rodentia, Mammalia) on the Island of Cyprus. Zootoxa *1241*, 1-36.

Cucchi, T., Vigne, J.D., and Auffray, J.C. (2005). First occurrence of the house mouse (Mus musculus domesticus Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences. Biological Journal of the Linnean Society *84*, 429 - 445.

Dalma-Weiszhausz, D.D., Warrington, J., Tanimoto, E.Y., and Miyada, C.G. (2006). The affymetrix GeneChip platform: an overview. Methods Enzymol *410*, 3-28.

Darwin, C. (1859). On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life (London, John Murray).

Davidson, E.H. (2001). Genomic regulatory systems (San Diego, Academic Press).

Dean, M.D., and Nachman, M.W. (2008). Faster Fertilization Rate in Conspecific Versus Heterospecific Matings in House Mice. Evolution.

Deng, Y., Ren, X., Yang, L., Lin, Y., and Wu, X. (2003). A JNK-dependent pathway is required for TNFalpha-induced apoptosis. Cell *115*, 61-70.

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature *330*, 576-578.

Dong, C., Yang, D.D., Tournier, C., Whitmarsh, A.J., Xu, J., Davis, R.J., and Flavell, R.A. (2000). JNK is required for effector T-cell function but not for T-cell activation. Nature *405*, 91-94.

Ellegren, H., and Sheldon, B.C. (2008). Genetic basis of fitness differences in natural populations. Nature *452*, 169-175.

Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S., Wiebe, V., Kitano, T., Monaco, A.P., and Paabo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. Nature *418*, 869-872.

English, J., Pearson, G., Wilsbacher, J., Swantek, J., Karandikar, M., Xu, S., and Cobb, M.H. (1999). New insights into the control of MAP kinase pathways. Exp Cell Res *253*, 255-270.

Evan, G.I., Lewis, G.K., Ramsay, G., and Bishop, J.M. (1985). Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. Mol Cell Biol *5*, 3610-3616.

Fleming, Y., Armstrong, C.G., Morrice, N., Paterson, A., Goedert, M., and Cohen, P. (2000). Synergistic activation of stress-activated protein kinase 1/c-Jun N-terminal kinase (SAPK1/JNK) isoforms by mitogen-activated protein kinase kinase 4 (MKK4) and MKK7. Biochem J *352 Pt 1*, 145-154.

Foltz, I.N., Gerl, R.E., Wieler, J.S., Luckach, M., Salmon, R.A., and Schrader, J.W. (1998). Human mitogen-activated protein kinase kinase 7 (MKK7) is a highly conserved c-Jun N-terminal kinase/stress-activated protein kinase (JNK/SAPK) activated by environmental stresses and physiological stimuli. J Biol Chem *273*, 9344-9351.

Forejt, J. (1996). Hybrid sterility in the mouse. Trends Genet *12*, 412-417.

Frazer, K.A., Eskin, E., Kang, H.M., Bogue, M.A., Hinds, D.A., Beilharz, E.J., Gupta, R.V., Montgomery, J., Morenzoni, M.M., Nilsen, G.B., *et al.* (2007). A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. Nature *448*, 1050-1053.

Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci U S A *85*, 8998-9002.

Fu, Y.X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics *147*, 915-925.

Funderburg, N., Lederman, M.M., Feng, Z., Drage, M.G., Jadlowsky, J., Harding, C.V., Weinberg, A., and Sieg, S.F. (2007). Human -defensin-3 activates professional antigen-presenting cells via Toll-like receptors 1 and 2. Proc Natl Acad Sci U S A *104*, 18631-18635.

Garner, D.L., and Johnson, L.A. (1995). Viability assessment of mammalian sperm using SYBR-14 and propidium iodide. Biol Reprod *53*, 276-284.

Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics *20*, 307-315.

Gordon, J.W., Scangos, G.A., Plotkin, D.J., Barbosa, J.A., and Ruddle, F.H. (1980). Genetic transformation of mouse embryos by microinjection of purified DNA. Proc Natl Acad Sci U S A *77*, 7380-7384.

Graber, J.H., Cantor, C.R., Mohr, S.C., and Smith, T.F. (1999). In silico detection of control signals: mRNA 3'-end-processing sequences in diverse species. Proc Natl Acad Sci U S A *96*, 14055-14060.

Graham, F.L., Smiley, J., Russell, W.C., and Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Virol *36*, 59-74.

Grosso, A.R., Gomes, A.Q., Barbosa-Morais, N.L., Caldeira, S., Thorne, N.P., Grech, G., von Lindern, M., and Carmo-Fonseca, M. (2008). Tissue-specific splicing factor gene expression signatures. Nucleic Acids Res *36*, 4823-4832.

Gu, H., Zou, Y.R., and Rajewsky, K. (1993). Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. Cell *73*, 1155-1164.

Guenet, J.L., and Bonhomme, F. (2003). Wild mice: an ever-increasing contribution to a popular mammalian model. Trends Genet *19*, 24-31.

Hald, J., Hjorth, J.P., German, M.S., Madsen, O.D., Serup, P., and Jensen, J. (2003). Activated Notch1 prevents differentiation of pancreatic acinar cells and attenuate endocrine development. Dev Biol *260*, 426-437.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser *41*, 95-98.

Han, S., Xie, W., Kim, S.H., Yue, L., and DeJong, J. (2004). A short core promoter drives expression of the ALF transcription factor in reproductive tissues of male and female mice. Biol Reprod *71*, 933-941.

Hanchard, N., Elzein, A., Trafford, C., Rockett, K., Pinder, M., Jallow, M., Harding, R., Kwiatkowski, D., and McKenzie, C. (2007). Classical sickle beta-globin haplotypes exhibit a high degree of long-range haplotype similarity in African and Afro-Caribbean populations. BMC Genet *8*, 52.

Harr, B., Voolstra, C., Heinen, T.J., Baines, J.F., Rottscheidt, R., Ihle, S., Muller, W., Bonhomme, F., and Tautz, D. (2006). A change of expression in the conserved signaling gene MKK7 is associated with a selective sweep in the western house mouse Mus musculus domesticus. J Evol Biol *19*, 1486-1496.

Hasler, J., Samuelsson, T., and Strub, K. (2007). Useful 'junk': Alu RNAs in the human transcriptome. Cell Mol Life Sci *64*, 1793-1800.

Heber, S., and Sick, B. (2006). Quality assessment of Affymetrix GeneChip data. OMICS *10*, 358-368.

Hilpert, K., Hansen, G., Wessner, H., Kuttner, G., Welfle, K., Seifert, M., and Hohne, W. (2001). Anti-c-myc antibody 9E10: epitope key positions and variability characterized using peptide spot synthesis on cellulose. Protein Eng *14*, 803-806.

Ho, D.T., Bardwell, A.J., Grewal, S., Iverson, C., and Bardwell, L. (2006). Interacting JNK-docking sites in MKK7 promote binding and activation of JNK mitogen-activated protein kinases. J Biol Chem *281*, 13169-13179.

Hobolth, A., Christensen, O.F., Mailund, T., and Schierup, M.H. (2007). Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. PLoS Genet *3*, e7.

Hoekstra, H.E., and Coyne, J.A. (2007). The locus of evolution: evo devo and the genetics of adaptation. Evolution *61*, 995-1016.

Holland, P.M., Suzanne, M., Campbell, J.S., Noselli, S., and Cooper, J.A. (1997). MKK7 is a stress-activated mitogen-activated protein kinase kinase functionally related to hemipterous. J Biol Chem *272*, 24994-24998.

Hu, J., Lutz, C.S., Wilusz, J., and Tian, B. (2005). Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. RNA *11*, 1485-1493.

Hunt, W.G., and Selander, R.K. (1973). Biochemical genetics of hybridisation in European house mice. Heredity *31*, 11-33.

Ihle, S. (2003). Detecting genes involved in selective sweeps within populations of the house mouse species complex - a multi-locus candidate gene approach. In Mathematisch-Naturwissenschaftliche Fakultät (Cologne, Universität zu Köln).

Ihle, S., Ravaoarimanana, I., Thomas, M., and Tautz, D. (2006). An analysis of signatures of selective sweeps in natural populations of the house mouse. Mol Biol Evol *23*, 790-797.

Imade, G.E., Towobola, O.A., Sagay, A.S., and Otubu, J.A. (1993). Discrepancies in sperm count using improved Neubauer, Makler, and Horwells counting chambers. Arch Androl *31*, 17-22.

Jansa, P., Divina, P., and Forejt, J. (2005). Construction and characterization of a genomic BAC library for the Mus m. musculus mouse subspecies (PWD/Ph inbred strain). BMC Genomics *6*, 161.

Kaiser, V., and Diamond, G. (2000). Expression of mammalian defensin genes. J Leukoc Biol *68*, 779-784.

Kauer, M.O., Dieringer, D., and Schlotterer, C. (2003). A microsatellite variability screen for positive selection associated with the "out of Africa" habitat expansion of Drosophila melanogaster. Genetics *165*, 1137-1148.

Kim, Y., and Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics *160*, 765-777.

Kimura, M. (1979). The neutral theory of molecular evolution. Sci Am *241*, 98-100, 102, 108 passim.

Kishimoto, H., Nakagawa, K., Watanabe, T., Kitagawa, D., Momose, H., Seo, J., Nishitai, G., Shimizu, N., Ohata, S., Tanemura, S., *et al.* (2003). Different properties of SEK1 and MKK7 in dual phosphorylation of stress-induced activated protein kinase SAPK/JNK in embryonic stem cells. J Biol Chem *278*, 16595-16601.

Kleene, K.C. (2001). A possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells. Mech Dev *106*, 3-23.

Kohany, O., Gentles, A.J., Hankus, L., and Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics *7*, 474.

Kohn, M.H., Pelz, H.J., and Wayne, R.K. (2000). Natural selection mapping of the warfarin-resistance gene. Proc Natl Acad Sci U S A *97*, 7911-7915.

Kohn, M.H., Pelz, H.J., and Wayne, R.K. (2003). Locus-specific genetic differentiation at Rw among warfarin-resistant rat (Rattus norvegicus) populations. Genetics *164*, 1055-1070.

Kos, C.H. (2004). Cre/loxP system for generating tissue-specific knockout mouse models. Nutr Rev *62*, 243-246.

Kozak, M. (1984). Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. Nature *308*, 241-246.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell *44*, 283-292.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res *15*, 8125-8148.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Lee, J.Y., Yeh, I., Park, J.Y., and Tian, B. (2007). PolyA_DB 2: mRNA polyadenylation sites in vertebrate genes. Nucleic Acids Res *35*, D165-168.

Lei, K., Nimnual, A., Zong, W.X., Kennedy, N.J., Flavell, R.A., Thompson, C.B., Bar-Sagi, D., and Davis, R.J. (2002). The Bax subfamily of Bcl2-related proteins is essential for apoptotic signal transduction by c-Jun NH(2)-terminal kinase. Mol Cell Biol *22*, 4929-4942.

Levine, M.T., Jones, C.D., Kern, A.D., Lindfors, H.A., and Begun, D.J. (2006). Novel genes derived from noncoding DNA in Drosophila melanogaster are frequently X-linked and exhibit testis-biased expression. Proc Natl Acad Sci U S A *103*, 9935-9939.

Li, S., Zhou, W., Doglio, L., and Goldberg, E. (1998). Transgenic mice demonstrate a testis-specific promoter for lactate dehydrogenase, LDHC. J Biol Chem *273*, 31191-31194.

Liu, D., Brockman, J.M., Dass, B., Hutchins, L.N., Singh, P., McCarrey, J.R., MacDonald, C.C., and Graber, J.H. (2007). Systematic variation in mRNA 3'-processing signals during mouse spermatogenesis. Nucleic Acids Res *35*, 234-246.

Liu, G., Loraine, A.E., Shigeta, R., Cline, M., Cheng, J., Valmeekam, V., Sun, S., Kulp, D., and Siani-Rose, M.A. (2003). NetAffx: Affymetrix probesets and annotations. Nucleic Acids Res *31*, 82-86.

Long, M., Betran, E., Thornton, K., and Wang, W. (2003). The origin of new genes: glimpses from the young and old. Nat Rev Genet *4*, 865-875.

Lynch, V.J., Tanzer, A., Wang, Y., Leung, F.C., Gellersen, B., Emera, D., and Wagner, G.P. (2008). Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals. Proc Natl Acad Sci U S A *105*, 14928-14933.

Macholan, M., Munclinger, P., Sugerkova, M., Dufkova, P., Bimova, B., Bozikova, E., Zima, J., and Pialek, J. (2007a). Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. Evolution *61*, 746-771.

Macholan, M., Vyskocilova, M., Bonhomme, F., Krystufek, B., Orth, A., and Vohralik, V. (2007b). Genetic variation and phylogeography of free-living mouse species (genus Mus) in the Balkans and the Middle East. Mol Ecol *16*, 4774-4788.

Makler, A. (1978). A new chamber for rapid sperm count and motility estimation. Fertil Steril *30*, 313-318.

Masui, S., Shimosato, D., Toyooka, Y., Yagi, R., Takahashi, K., and Niwa, H. (2005). An efficient system to establish multiple embryonic stem cell lines carrying an inducible expression unit. Nucleic Acids Res *33*, e43.

Maxwell, A.I., Morrison, G.M., and Dorin, J.R. (2003). Rapid sequence divergence in mammalian beta-defensins by adaptive evolution. Mol Immunol *40*, 413-421.

McGregor, A.P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D.G., Payre, F., and Stern, D.L. (2007). Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature *448*, 587-590.

Metta, M., and Schlotterer, C. (2008). Male-biased genes are overrepresented among novel Drosophila pseudoobscura sex-biased genes. BMC Evol Biol *8*, 182.

Moriguchi, T., Toyoshima, F., Masuyama, N., Hanafusa, H., Gotoh, Y., and Nishida, E. (1997). A novel SAPK/JNK kinase, MKK7, stimulated by TNFalpha and cellular stresses. EMBO J *16*, 7045-7053.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol *51 Pt 1*, 263-273.

Muyrers, J.P., Zhang, Y., Testa, G., and Stewart, A.F. (1999). Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res *27*, 1555-1557.

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W., and Roder, J.C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. Proc Natl Acad Sci U S A *90*, 8424-8428.

Nishina, H., Wada, T., and Katada, T. (2004). Physiological roles of SAPK/JNK signaling pathway. J Biochem *136*, 123-126.

Ohno, S. (1970). Evolution by gene duplication (New York, Springer Verlag).

Ohta, T. (1988). Evolution by gene duplication and compensatory advantageous mutations. Genetics *120*, 841-847.

Okabe, S. (2001). Gene expression in transgenic mice using neural promoters. Curr Protoc Neurosci *Chapter 3*, Unit 3 16.

Osoegawa, K., Tateno, M., Woon, P.Y., Frengen, E., Mammoser, A.G., Catanese, J.J., Hayashizaki, Y., and de Jong, P.J. (2000). Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. Genome Res *10*, 116-128.

Otte, J.M., Werner, I., Brand, S., Chromik, A.M., Schmitz, F., Kleine, M., and Schmidt, W.E. (2008). Human beta defensin 2 promotes intestinal wound healing in vitro. J Cell Biochem *104*, 2286-2297.

Otto-Duessel, M., Khankaldyyan, V., Gonzalez-Gomez, I., Jensen, M.C., Laug, W.E., and Rosol, M. (2006). In vivo testing of Renilla luciferase substrate analogs in an orthotopic murine model of human glioblastoma. Mol Imaging *5*, 57-64.

Patil, A.A., Cai, Y., Sang, Y., Blecha, F., and Zhang, G. (2005). Cross-species analysis of the mammalian beta-defensin gene family: presence of syntenic gene clusters and preferential expression in the male reproductive tract. Physiol Genomics *23*, 5-17.

Pelz, H.J., Rost, S., Hunerberg, M., Fregin, A., Heiberg, A.C., Baert, K., MacNicoll, A.D., Prescott, C.V., Walker, A.S., Oldenburg, J., *et al.* (2005). The genetic basis of resistance to anticoagulants in rodents. Genetics *170*, 1839-1847.

Perkinton, M.S., Standen, C.L., Lau, K.F., Kesavapany, S., Byers, H.L., Ward, M., McLoughlin, D.M., and Miller, C.C. (2004). The c-Abl tyrosine kinase phosphorylates the Fe65 adaptor protein to stimulate Fe65/amyloid precursor protein nuclear signaling. J Biol Chem *279*, 22084-22091.

Proudfoot, N., and O'Sullivan, J. (2002). Polyadenylation: a tail of two complexes. Curr Biol *12*, R855-857.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501-512.

Rajewsky, K., Gu, H., Kuhn, R., Betz, U.A., Muller, W., Roes, J., and Schwenk, F. (1996). Conditional gene targeting. J Clin Invest *98*, 600-603.

Ray, P., and Gambhir, S.S. (2007). Noninvasive imaging of molecular events with bioluminescent reporter genes in living subjects. Methods Mol Biol *411*, 131-144.

Reddi, P.P., Flickinger, C.J., and Herr, J.C. (1999). Round spermatid-specific transcription of the mouse SP-10 gene is mediated by a 294-base pair proximal promoter. Biol Reprod *61*, 1256-1266.

Reddi, P.P., Shore, A.N., Shapiro, J.A., Anderson, A., Stoler, M.H., and Acharya, K.K. (2003). Spermatid-specific promoter of the SP-10 gene functions as an insulator in somatic cells. Dev Biol *262*, 173-182.

Reddi, P.P., Urekar, C.J., Abhyankar, M.M., and Ranpura, S.A. (2007). Role of an insulator in testis-specific gene transcription. Ann N Y Acad Sci *1120*, 95-103.

Rockman, M.V., and Stern, D.L. (2008). Tinker where the tinkering's good. Trends Genet *24*, 317-319.

Rottscheidt, R. (2007). Gene Expression Divergence between Subspecies of the House Mouse and the Contribution to Reproductive Isolation. In Mathematisch-Naturwissenschaftliche Fakultät (Cologne, Universität zu Köln).

Rottscheidt, R., and Harr, B. (2007). Extensive additivity of gene expression differentiates subspecies of the house mouse. Genetics *177*, 1553-1567.

Russell, L.D., Ettlin, R.A., Sinha Hikim, A.P., and Clegg, E.D. (1990). Histological and histopathological evaluation of the testis (St. Louis, Cache River Press).

Sabapathy, K., Jochum, W., Hochedlinger, K., Chang, L., Karin, M., and Wagner, E.F. (1999). Defective neural tube morphogenesis and altered apoptosis in the absence of both JNK1 and JNK2. Mech Dev *89*, 115-124.

Sadikot, R.T., and Blackwell, T.S. (2005). Bioluminescence imaging. Proc Am Thorac Soc *2*, 537-540, 511-532.

Sage, R.D., Heyneman, D., Lim, K.C., and Wilson, A.C. (1986). Wormy mice in a hybrid zone. Nature *324*, 60-63.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York, Cold Sprong Harbor Press).

Sauer, B. (2002). Cre/lox: one more step in the taming of the genome. Endocrine *19*, 221-228.

Saunders, T.L. (2003). Reporter molecules in genetically engineered mice. Methods Mol Biol *209*, 125-143.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467-470.

Schonig, K., and Bujard, H. (2003). Generating conditional mouse mutants via tetracycline-controlled gene expression. Methods Mol Biol *209*, 69-104.

Schultz, N., Hamra, F.K., and Garbers, D.L. (2003). A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. Proc Natl Acad Sci U S A *100*, 12201-12206.

Schwenk, F., Baron, U., and Rajewsky, K. (1995). A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. Nucleic Acids Res *23*, 5080-5081.

Scieglinska, D., Vydra, N., Krawczyk, Z., and Widlak, W. (2004). Location of promoter elements necessary and sufficient to direct testis-specific expression of the Hst70/Hsp70.2 gene. Biochem J *379*, 739-747.

Semple, C.A., Taylor, K., Eastwood, H., Barran, P.E., and Dorin, J.R. (2006). Beta-defensin evolution: selection complexity and clues for residues of functional importance. Biochem Soc Trans *34*, 257-262.

Shiffman, D., and Porter, J.G. (2000). Gene expression profiling of cardiovascular disease models. Curr Opin Biotechnol *11*, 598-601.

Slott, V.L., Suarez, J.D., Poss, P.M., Linder, R.E., Strader, L.F., and Perreault, S.D. (1993). Optimization of the Hamilton-Thorn computerized sperm motility analysis

system for use with rat spermatozoa in toxicological studies. Fundam Appl Toxicol *21*, 298-307.

Smith, J.M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. Genet Res *23*, 23-35.

Somboonthum, P., Ohta, H., Yamada, S., Onishi, M., Ike, A., Nishimune, Y., and Nozaki, M. (2005). cAMP-responsive element in TATA-less core promoter is essential for haploid-specific gene expression in mouse testis. Nucleic Acids Res *33*, 3401-3411.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol *98*, 503-517.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123*, 585-595.

Takekawa, M., Tatebayashi, K., and Saito, H. (2005). Conserved docking site is essential for activation of mammalian MAP kinase kinases by specific MAP kinase kinase kinases. Mol Cell *18*, 295-306.

Tautz, D., and Pfeifle, C. (1989). A non-radioactive in situ hybridization method for the localization of specific RNAs in Drosophila embryos reveals translational control of the segmentation gene hunchback. Chromosoma *98*, 81-85.

Teeter, K.C., Payseur, B.A., Harris, L.W., Bakewell, M.A., Thibodeau, L.M., O'Brien, J.E., Krenz, J.G., Sans-Fuentes, M.A., Nachman, M.W., and Tucker, P.K. (2008). Genome-wide patterns of gene flow across a house mouse hybrid zone. Genome Res *18*, 67-76.

Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503-512.

Todaro, G.J., and Green, H. (1963). Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. J Cell Biol *17*, 299-313.

Tollner, T.L., Yudin, A.I., Treece, C.A., Overstreet, J.W., and Cherr, G.N. (2008). Macaque sperm coating protein DEFB126 facilitates sperm penetration of cervical mucus. Hum Reprod.

Topaloglu, O., Schluter, G., Nayernia, K., and Engel, W. (2001). A 74-bp promoter of the Tnp2 gene confers testis- and spermatid-specific expression in transgenic mice. Biochem Biophys Res Commun *289*, 597-601.

Torres, T.T., Metta, M., Ottenwalder, B., and Schlotterer, C. (2008). Gene expression profiling by massively parallel sequencing. Genome Res *18*, 172-177.

Tournier, C., Dong, C., Turner, T.K., Jones, S.N., Flavell, R.A., and Davis, R.J. (2001). MKK7 is an essential component of the JNK signal transduction pathway activated by proinflammatory cytokines. Genes Dev *15*, 1419-1426.

Tournier, C., Whitmarsh, A.J., Cavanagh, J., Barrett, T., and Davis, R.J. (1997). Mitogen-activated protein kinase kinase 7 is an activator of the c-Jun NH2-terminal kinase. Proc Natl Acad Sci U S A *94*, 7337-7342.

Tournier, C., Whitmarsh, A.J., Cavanagh, J., Barrett, T., and Davis, R.J. (1999). The MKK7 gene encodes a group of c-Jun NH2-terminal kinase kinases. Mol Cell Biol *19*, 1569-1581.

Tucker, P.K., Sage, R.D., Warner, J.H., Wilson, A.C., and Eicher, E.M. (1992). Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. Evolution *46*, 1146-1163.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A *98*, 5116-5121.

Valenzuela, D.M., Murphy, A.J., Frendewey, D., Gale, N.W., Economides, A.N., Auerbach, W., Poueymirou, W.T., Adams, N.C., Rojas, J., Yasenchak, J., *et al.* (2003). High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. Nat Biotechnol *21*, 652-659.

Venkataraman, K., Brown, K.M., and Gilmartin, G.M. (2005). Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes Dev *19*, 1315-1327.

Volloch, V., Schweitzer, B., and Rits, S. (1994). Ligation-mediated amplification of RNA from murine erythroid cells reveals a novel class of beta globin mRNA with an extended 5'-untranslated region. Nucleic Acids Res *22*, 2507-2511.

Voolstra, C., Tautz, D., Farbrother, P., Eichinger, L., and Harr, B. (2007). Contrasting evolution of expression differences in the testis between species and subspecies of the house mouse. Genome Res *17*, 42-49.

Wada, T., Joza, N., Cheng, H.Y., Sasaki, T., Kozieradzki, I., Bachmaier, K., Katada, T., Schreiber, M., Wagner, E.F., Nishina, H., *et al.* (2004). MKK7 couples stress signalling to G2/M cell-cycle progression and cellular senescence. Nat Cell Biol *6*, 215-226.

Wagner, G.P., and Lynch, V.J. (2008). The gene regulatory logic of transcription factor evolution. Trends Ecol Evol *23*, 377-385.

Wang, X., Destrument, A., and Tournier, C. (2007). Physiological roles of MKK4 and MKK7: insights from animal models. Biochim Biophys Acta *1773*, 1349-1357.

Wiehe, T., Nolte, V., Zivkovic, D., and Schlotterer, C. (2007). Identification of selective sweeps using a dynamically adjusted number of linked microsatellites. Genetics *175*, 207-218.

Winter, J., Kunath, M., Roepcke, S., Krause, S., Schneider, R., and Schweiger, S. (2007). Alternative polyadenylation signals and promoters act in concert to control tissue-specific expression of the Opitz Syndrome gene MID1. BMC Mol Biol *8*, 105.

World Health Organization (1992). WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction (Cambridge, Cambridge University Press).

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat Rev Genet *8*, 206-216.

Yaffe, D., and Saxel, O. (1977). A myogenic cell line with altered serum requirements for differentiation. Differentiation *7*, 159-166.

Yamaguchi, Y., Fukuhara, S., Nagase, T., Tomita, T., Hitomi, S., Kimura, S., Kurihara, H., and Ouchi, Y. (2001). A novel mouse beta-defensin, mBD-6, predominantly expressed in skeletal muscle. J Biol Chem *276*, 31510-31514.

Yamaguchi, Y., Nagase, T., Tomita, T., Nakamura, K., Fukuhara, S., Amano, T., Yamamoto, H., Ide, Y., Suzuki, M., Teramoto, S., *et al.* (2007). Beta-defensin overexpression induces progressive muscle degeneration in mice. Am J Physiol Cell Physiol *292*, C2141-2149.

Yang, L.V., Heng, H.H., Wan, J., Southwood, C.M., Gow, A., and Li, L. (2003a). Alternative promoters and polyadenylation regulate tissue-specific expression of Hemogen isoforms during hematopoiesis and spermatogenesis. Dev Dyn *228*, 606-616.

Yang, S.H., Sharrocks, A.D., and Whitmarsh, A.J. (2003b). Transcriptional regulation by the MAP kinase signaling cascades. Gene *320*, 3-21.

Yang, X.W., Model, P., and Heintz, N. (1997). Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. Nat Biotechnol *15*, 859-865.

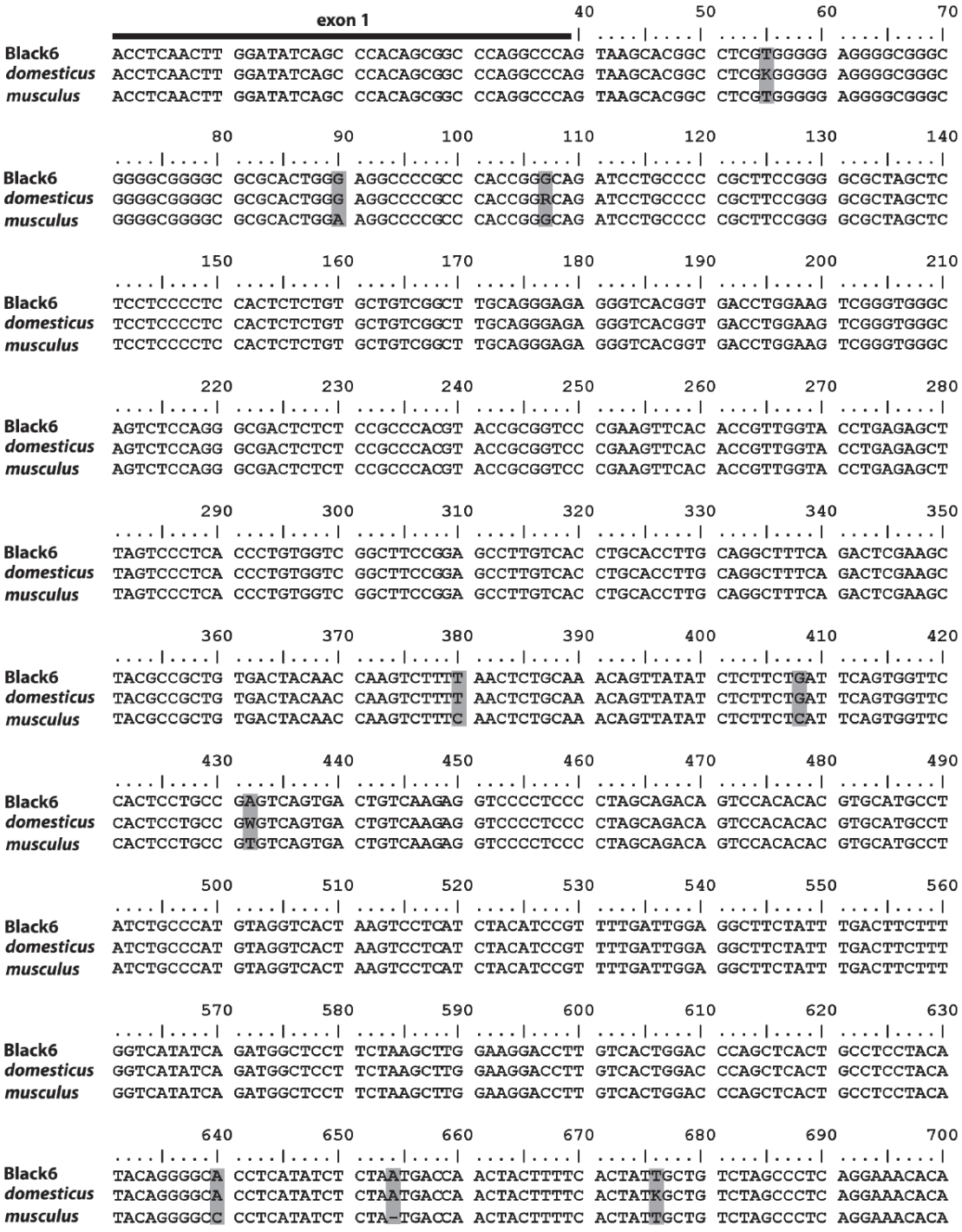
Zambrowicz, B.P., Harendza, C.J., Zimmermann, J.W., Brinster, R.L., and Palmiter, R.D. (1993). Analysis of the mouse protamine 1 promoter in transgenic mice. Proc Natl Acad Sci U S A *90*, 5071-5075.

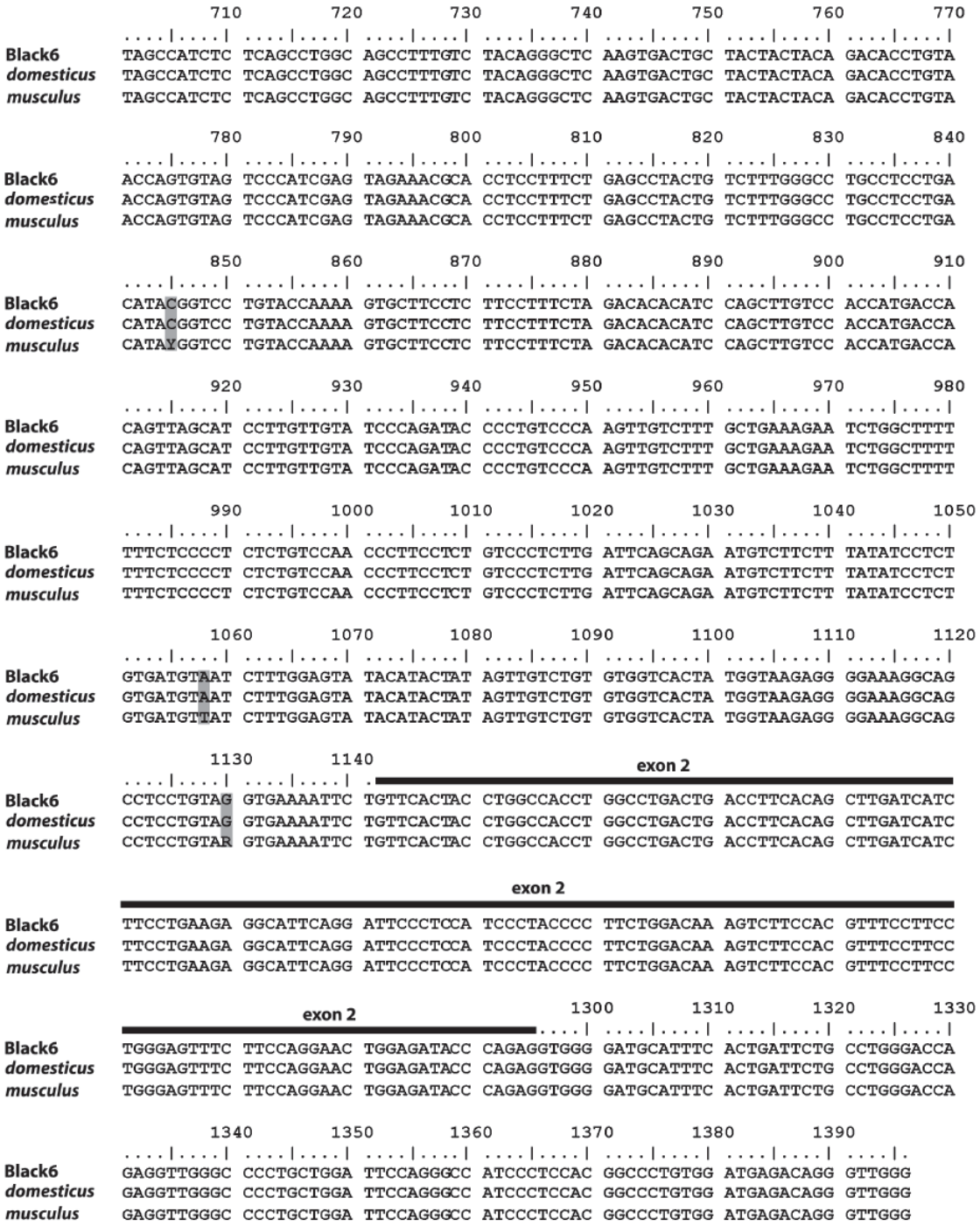
Zhou, Q., Zhang, G., Zhang, Y., Xu, S., Zhao, R., Zhan, Z., Li, X., Ding, Y., Yang, S., and Wang, W. (2008). On the origin of new genes in Drosophila. Genome Res *18*, 1446-1455.

Zinn, K.R., Chaudhuri, T.R., Szafran, A.A., O'Quinn, D., Weaver, C., Dugger, K., Lamar, D., Kesterson, R.A., Wang, X., and Frank, S.J. (2008). Noninvasive bioluminescence imaging in small animals. ILAR J *49*, 103-115.

9 Supplement A

Genomic sequence alignment of Mkk7 intron 1 and flanking parts. Consensus sequences of 8 animals each from *Mus musculus musculus* and *Mus musculus domesticus* are aligned against Black6 sequence. Relative base pair positions are indicated. SNP positions are highlighted with grey background color. Exon positions are indicated.




710 720 730 740 750 760 770
Black6 TAGCCATCTC TCAGCCTGGC AGCCTTTGTC TACAGGGCTC AAGTGACTGC TACTACTACA GACACCTGTA
domesticus TAGCCATCTC TCAGCCTGGC AGCCTTTGTC TACAGGGCTC AAGTGACTGC TACTACTACA GACACCTGTA
musculus TAGCCATCTC TCAGCCTGGC AGCCTTTGTC TACAGGGCTC AAGTGACTGC TACTACTACA GACACCTGTA
780 790 800 810 820 830 840
Black6 ACCAGTGTAG TCCCATCGAG TAGAAACGCA CCTCCTTTCT GAGCCTACTG TCTTTGGGCC TGCCTCCTGA
domesticus ACCAGTGTAG TCCCATCGAG TAGAAACGCA CCTCCTTTCT GAGCCTACTG TCTTTGGGCC TGCCTCCTGA
musculus ACCAGTGTAG TCCCATCGAG TAGAAACGCA CCTCCTTTCT GAGCCTACTG TCTTTGGGCC TGCCTCCTGA
850 860 870 880 890 900 910
Black6 CATACGGTCC TGTACCAAAA GTGCTTCCTC TTCCTTTCTA GACACACATC CAGCTTGTCC ACCATGACCA
domesticus CATACGGTCC TGTACCAAAA GTGCTTCCTC TTCCTTTCTA GACACACATC CAGCTTGTCC ACCATGACCA
musculus CATAYGGTCC TGTACCAAAA GTGCTTCCTC TTCCTTTCTA GACACACATC CAGCTTGTCC ACCATGACCA
920 930 940 950 960 970 980
Black6 CAGTTAGCAT CCTTGTTGTA TCCCAGATAC CCCTGTCCCA AGTTGTCTTT GCTGAAAGAA TCTGGCTTTT
domesticus CAGTTAGCAT CCTTGTTGTA TCCCAGATAC CCCTGTCCCA AGTTGTCTTT GCTGAAAGAA TCTGGCTTTT
musculus CAGTTAGCAT CCTTGTTGTA TCCCAGATAC CCCTGTCCCA AGTTGTCTTT GCTGAAAGAA TCTGGCTTTT
990 1000 1010 1020 1030 1040 1050
Black6 TTTCTCCCCT CTCTGTCCAA CCCTTCCTCT GTCCCTCTTG ATTCAGCAGA ATGTCTTCTT TATATCCTCT
domesticus TTTCTCCCCT CTCTGTCCAA CCCTTCCTCT GTCCCTCTTG ATTCAGCAGA ATGTCTTCTT TATATCCTCT
musculus TTTCTCCCCT CTCTGTCCAA CCCTTCCTCT GTCCCTCTTG ATTCAGCAGA ATGTCTTCTT TATATCCTCT
1060 1070 1080 1090 1100 1110 1120
Black6 GTGATGTAAT CTTTGGAGTA TACATACTAT AGTTGTCTGT GTGGTCACTA TGGTAAGAGG GGAAAGGCAG
domesticus GTGATGTAAT CTTTGGAGTA TACATACTAT AGTTGTCTGT GTGGTCACTA TGGTAAGAGG GGAAAGGCAG
musculus GTGATGTTAT CTTTGGAGTA TACATACTAT AGTTGTCTGT GTGGTCACTA TGGTAAGAGG GGAAAGGCAG
1130 1140
exon 2
Black6 CCTCCTGTAG GTGAAAATTC TGTTCACTAC CTGGCCACCT GGCCTGACTG ACCTTCACAG CTTGATCATC
domesticus CCTCCTGTAG GTGAAAATTC TGTTCACTAC CTGGCCACCT GGCCTGACTG ACCTTCACAG CTTGATCATC
musculus CCTCCTGTAR GTGAAAATTC TGTTCACTAC CTGGCCACCT GGCCTGACTG ACCTTCACAG CTTGATCATC
exon 2
Black6 TTCCTGAAGA GGCATTCAGG ATTCCCTCCA TCCCTACCCC TTCTGGACAA AGTCTTCCAC GTTTCCTTCC
domesticus TTCCTGAAGA GGCATTCAGG ATTCCCTCCA TCCCTACCCC TTCTGGACAA AGTCTTCCAC GTTTCCTTCC
musculus TTCCTGAAGA GGCATTCAGG ATTCCCTCCA TCCCTACCCC TTCTGGACAA AGTCTTCCAC GTTTCCTTCC
exon 2
1300 1310 1320 1330
Black6 TGGGAGTTTC TTCCAGGAAC TGGAGATACC CAGAGGTGGG GATGCATTTC ACTGATTCTG CCTGGGACCA
domesticus TGGGAGTTTC TTCCAGGAAC TGGAGATACC CAGAGGTGGG GATGCATTTC ACTGATTCTG CCTGGGACCA
musculus TGGGAGTTTC TTCCAGGAAC TGGAGATACC CAGAGGTGGG GATGCATTTC ACTGATTCTG CCTGGGACCA
1340 1350 1360 1370 1380 1390
Black6 GAGGTTGGGC CCCTGCTGGA TTCCAGGGCC ATCCCTCCAC GGCCCTGTGG ATGAGACAGG GTTGGG
domesticus GAGGTTGGGC CCCTGCTGGA TTCCAGGGCC ATCCCTCCAC GGCCCTGTGG ATGAGACAGG GTTGGG
musculus GAGGTTGGGC CCCTGCTGGA TTCCAGGGCC ATCCCTCCAC GGCCCTGTGG ATGAGACAGG GTTGGG

10 Supplement B

Results of the Affymetrix Genechip experiment comparing testis expression in LP10/Hennes knock-out mice and wild-type mice. Significantly differentially expressed genes calculated by SAM analysis (FDR=4%, delta=1.98) are presented in SAM output format.

higher expressed in wild type testis						
Gene ID	Gene Name	Score (d)	Numerator (r)	Denominator (s+s0)	Fold Change	q-value (%)
1432460_at	1700125F08Rik	24,80	1076,62	43,41	14,05	0,00
1452534_a_at	Hmgb2	9,94	488,86	49,20	1,16	0,00
1451847_s_at	Arid4b	8,18	107,34	13,13	1,64	2,60
1436886_x_at	Xab2	8,09	179,96	22,26	1,36	2,60

higher expressed in LP10/Hennes knock-out testis						
Gene ID	Gene Name	Score (d)	Numerator (r)	Denominator (s+s0)	Fold Change	q-value (%)
1419451_at	Fzr1	-16,29	-322,03	19,77	0,86	0,00
1436143_at	4933425L03Rik	-11,47	-444,66	38,75	0,82	0,00
1456823_at	Gm70	-10,44	-739,44	70,80	0,85	0,00
1423444_at	Rock1	-10,24	-311,89	30,45	0,86	0,00
1434782_at	Usp42	-9,91	-1365,43	137,84	0,78	0,00
1423293_at	Rpa1	-9,38	-617,01	65,79	0,88	0,00
1417684_at	Thumpd3	-8,99	-491,80	54,68	0,90	0,00
1417425_at	Prkrip1	-8,76	-172,12	19,65	0,88	0,00
1460000_at	Shisa3	-8,66	-132,23	15,27	0,86	0,00
1423553_at	Dnajb3	-8,58	-1666,54	194,24	0,89	0,00
1418757_at	Trim69	-8,56	-602,46	70,35	0,88	0,00
1416868_at	Cdkn2c	-8,47	-361,79	42,73	0,88	0,00
1451470_s_at	Eif5a	-8,28	-1031,55	124,66	0,95	0,00
1448903_at	Sep15	-7,83	-556,57	71,09	0,88	0,00
1417838_at	Ssty2	-7,52	-2549,77	339,13	0,88	0,00
1452040_a_at	Cdca3	-7,45	-990,17	132,94	0,89	0,00
1432186_at	1700028J19Rik	-7,43	-1325,48	178,42	0,87	0,00
1423801_a_at	Aprt	-7,24	-544,33	75,16	0,86	3,90
1424712_at	Ahctf1	-7,06	-1016,22	144,03	0,86	3,90
1418473_at	Cutc	-7,00	-386,83	55,29	0,80	3,90
1423833_a_at	Brp44	-6,91	-2732,58	395,66	0,87	3,90
1456660_a_at	0610010F05Rik	-6,84	-301,44	44,05	0,89	3,90
1430886_at	1700112E06Rik	-6,68	-171,27	25,64	0,06	3,90

1451337_at	Psmf1	-6,67	-1383,45	207,45	0,86	3,90
1420920_a_at	Arf1	-6,66	-337,16	50,63	0,90	3,90
1420851_at	Pard6g	-6,57	-109,90	16,73	0,77	3,90
1431086_s_at	Pcmt1	-6,34	-422,94	66,76	0,91	4,59
1442871_at	LOC100042492	-6,28	-486,48	77,53	0,54	4,59
1449177_at	Ccna1	-6,24	-556,35	89,17	0,81	4,59
1432353_at	Larp2	-6,21	-141,24	22,75	0,81	4,59
1454726_s_at	Ptpdc1	-6,07	-502,81	82,83	0,91	4,59
1452499_a_at	Kif2a	-6,05	-410,17	67,75	0,86	4,59
1418576_at	Yipf5	-5,97	-465,00	77,93	0,84	4,59
1431871_at	Txndc3	-5,95	-405,44	68,19	0,89	4,59
1416415_a_at	H2afz	-5,88	-2068,67	351,58	0,91	4,59
1422471_at	Pex13	-5,80	-941,83	162,30	0,89	5,57
1437535_at	Ppp3r2	-5,76	-1432,61	248,93	0,89	5,57

11 Supplement C

Gene ontology annotations of genes that are significantly higher expressed in testis of LP10/Hennes knock-out mice. Gene list is based on SAM-analysis shown in supplement B.

GOID	GO_term	Frequency	Genome frequency	Corrected P-value	Gene(s)
GO:0007126	Meiosis	0,05405	0,00171	0,0843	Fzr1,Rpa1,
GO:0051327	M phase of meiotic cell cycle	0,05405	0,00171	0,0843	Fzr1,Rpa1,
GO:0051321	meiotic cell cycle	0,05405	0,00173	0,0872	Fzr1,Rpa1,
GO:0000279	M phase	0,05405	0,00285	0,2311	Fzr1,Rpa1,
GO:0009987	cellular process	0,35135	0,16950	0,2705	Pex13,Cdkn2c,Sep15,Arf1,Ahctf1,Aprt,Rock1,Prkrip1,Ppp3r2,Fzr1,Eif5a,Pcmt1,Rpa1,
GO:0022403	cell cycle phase	0,05405	0,00400	0,4437	Fzr1,Rpa1,
GO:0044237	cellular metabolic process	0,18919	0,06957	0,5838	Aprt,Pex13,Prkrip1,Ppp3r2,Sep15,Ahctf1,Pcmt1,
GO:0022402	cell cycle process	0,05405	0,00494	0,6635	Fzr1,Rpa1,
GO:0008152	metabolic process	0,18919	0,07174	0,6832	Aprt,Pex13,Prkrip1,Ppp3r2,Sep15,Ahctf1,Pcmt1,
GO:0015031	protein transport	0,05405	0,00547	0,8038	Pex13,Arf1,
GO:0045184	establishment of protein localization	0,05405	0,00550	0,8120	Pex13,Arf1,
GO:0044238	primary metabolic process	0,16216	0,06483	1,0000	Aprt,Pex13,Ppp3r2,Sep15,Ahctf1,Pcmt1,
GO:0008104	protein localization	0,05405	0,00750	1,0000	Pex13,Arf1,
GO:0007049	cell cycle	0,05405	0,00753	1,0000	Fzr1,Rpa1,
GO:0007010	cytoskeleton organization and biogenesis	0,05405	0,00759	1,0000	Pex13,Rock1,
GO:0033036	macromolecule localization	0,05405	0,00785	1,0000	Pex13,Arf1,
GO:0007610	behavior	0,05405	0,00920	1,0000	Aprt,Pex13,
GO:0043170	macromolecule metabolic process	0,13514	0,05407	1,0000	Pex13,Ppp3r2,Sep15,Ahctf1,Pcmt1,
GO:0048523	negative regulation of cellular process	0,08108	0,02408	1,0000	Cdkn2c,Prkrip1,Rock1,
GO:0044260	cellular macromolecule metabolic process	0,08108	0,02602	1,0000	Ppp3r2,Sep15,Pcmt1,

GO:0048519	negative regulation of biological process	0,08108	0,02637	1,0000	Cdkn2c,Prkrip1, Rock1,
GO:0006915	apoptosis	0,05405	0,01426	1,0000	Rock1,Eif5a,
GO:0012501	programmed cell death	0,05405	0,01455	1,0000	Rock1,Eif5a,
GO:0008219	cell death	0,05405	0,01497	1,0000	Rock1,Eif5a,
GO:0016265	death	0,05405	0,01526	1,0000	Rock1,Eif5a,
GO:0006996	organelle organization and biogenesis	0,05405	0,01585	1,0000	Pex13,Rock1,
GO:0050896	response to stimulus	0,08108	0,03517	1,0000	Aprt,Pex13,Ppp 3r2,
GO:0044267	cellular protein metabolic process	0,05405	0,02255	1,0000	Sep15,Pcmt1,
GO:0019538	protein metabolic process	0,05405	0,02349	1,0000	Sep15,Pcmt1,
GO:0006810	transport	0,05405	0,02414	1,0000	Pex13,Arf1,
GO:0051234	establishment of localization	0,05405	0,02520	1,0000	Pex13,Arf1,
GO:0043283	biopolymer metabolic process	0,08108	0,04666	1,0000	Ppp3r2,Ahctf1,P cmt1,
GO:0031323	regulation of cellular metabolic process	0,05405	0,02876	1,0000	Prkrip1,Ahctf1,
GO:0019222	regulation of metabolic process	0,05405	0,02955	1,0000	Prkrip1,Ahctf1,
GO:0016043	cellular component organization and biogenesis	0,05405	0,02981	1,0000	Pex13,Rock1,
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0,05405	0,03284	1,0000	Aprt,Ahctf1,
GO:0048869	cellular developmental process	0,05405	0,03481	1,0000	Pex13,Rock1,
GO:0051179	localization	0,05405	0,03561	1,0000	Pex13,Arf1,
GO:0032502	developmental process	0,08108	0,06904	1,0000	Pex13,Rock1,Eif 5a,
GO:0050794	regulation of cellular process	0,10811	0,10294	1,0000	Cdkn2c,Prkrip1, Rock1,Ahctf1,
GO:0050789	regulation of biological process	0,10811	0,10649	1,0000	Cdkn2c,Prkrip1, Rock1,Ahctf1,
GO:0048856	anatomical structure development	0,05405	0,05425	1,0000	Pex13,Rock1,
GO:0065007	biological regulation	0,10811	0,11499	1,0000	Cdkn2c,Prkrip1, Rock1,Ahctf1,

GO:0032501	multicellular organismal process	0,05405	0,10411	1,0000	Pex13,Ppp3r2,
GO:XXXXX	unannotated	0,54054	0,75334	1,0000	0610010F05Rik,Usp42,Pard6g,Brp44,4933425L03Rik,Dnajb3,1700028J19Rik,H2afz,Gm70,Kif2a,1700112E06Rik,Txndc3,Ccna1,Trim69,Shisa3,Larp2,LOC100042492,Ptpdc1,Psmf1,Ssty2,
GO:0008150	biological_process	0,10811	0,73391	1,0000	0610010F05Rik,Pard6g,4933425L03Rik,1700028J19Rik,Dnajb3,H2afz,Ppp3r2,1700112E06Rik,Fzr1,Txndc3,Pcmt1,Trim69,Shisa3,Larp2,Cdkn2c,Sep15,Ahctf1,Psmf1,Rock1,Eif5a,Cutc,Yipf5,Usp42,Arf1,Thumpd3,Brp44,Prkrip1,Gm70,Kif2a,Cdca3,Ccna1,Pex13,LOC100042492,Aprt,Ptpdc

12 Supplement D

Genomic Sequence alignment of beta-Defensin6 promoter. Consensus sequences were built of 8 individuals from *M. m. domesticus* and *M. m. musculus* and aligned alongside Black6 database sequence (www.ENSEMBL.org). Identical sequences are illustrated as dashed line and mutations are indicated. Heterozygous indels are marked as "X". A 829 bp endogenous retrovirus is present in *M. m. musculus*. 3 out of the 16 haplotypes that were analyzed do not have the insertion. Those three alleles are similar to the *M. m. domesticus* type in terms of most SNPs and indels as well. All other alleles have *M. m. musculus* specific SNPs and indels shown by ambiguity code. The transcription start is indicated by highlighting the first exon with grey background. 17 nucleotide positions in the promoter upstream of the transcription start and the first exons are not included into the alignment, but they are known to contain no differences between *M. m. musculus* and *M. m. domesticus* (Sonja Ihle, personal communication).

```
Black6      ATAATAAAAA GATAGGAGAA TTATTTCTAG TCGATCTATT GCATGCTTAT ACCACTGAAG ATAATGAAAA
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      GCTGCTATTA GTATAGTAAA TTTATTCAAT TTGAAGGGAT GGATATACTA ATTATTCTGT ATTGATCCTT
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... W......... .......... .......... .......... ..........

Black6      GCCTACCATA TTTATATACT TTGATAGTAC ACAATAGTAC ACAAATATGT ATATTTACTG TGTCTTATTT
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... ..R....... ..M....... .......... ...Y......

Black6      TATTTTAAAA TCTCTTCTGT ATACCAGGTT AAAAATATGT CTACCAAGCT CAATACTGTG AGATTTTCCA
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      ACACAGCAGT GGGAGTGAGG CATTCTAATG TGGGTGCAAA CTCTTCATTC TCCAGCCTCA AATCTCTGTG
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      CTTCGGATTC TATAGAACAC CTTTTGTCAA AACTGTTTAT ACATTGTAGA TTTCTTCTTA TGAGGTTGTT
domesticus  .......... .......Y.. .......... .......... .......... .......... ..........
musculus    ...S...... .......Y.. .......... .......... .....R.... .......... ..W.......

Black6      TTTTTAATCT TCCTGTAGGG AGGAATAAAT CCTGGTTAAT GTAACTTTTT TTCACTTACA AAATCCTAGT
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    ....-..... .......... .......... .......... .......... .......... ..........

Black6      CTCTCAGCTC TTGCCATGAG GTATGCATTG ATAGTCATCA ATGGTTTTCT AGGATAAAAA TAAACAGCTG
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... ........R. .......... R......... .......... .......... ..........

Black6      GGTTATGGAA TTCACAAGAG ACCTCACTCA GAGTCAAGCC ACGTGGCCTC CCTCAGAACA GGTTTGAGCA
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      GGAGAAACAG GGAGATGGTG GAAAGGACTA GAGTCCCACT GTTATGGATT CTCATAAAAG ATTCTAAAAA
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      GCATCTTGAC TCCTCCCCTC TGATACTCAC ATTATTTCAG AGAGTGCAGA ACCACATATC CAAATGCTAT
domesticus  .......... .......... .......... .......... .R........ .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      CAAGTCCCAC ATTTGTCCAT CCCCTTGCGG CTAGAAGTAG CTATTTACTA GTTTTTTTTT CCTAAGGGAG
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........
```

```
Black6      AATAGTGCCC CATGTTCAGG TAACATGTTA CTAACACATT GCTTTCTCTG AGACTGAAAA CAACACACTC
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      ATCTGACTCC AGGTCCATCA CACTGAGTTC CTGGAGGTGT TTAGTAGTAA ATGAACTTCA ACATTGAGCA
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... ..Y....... .......... .......... .......... ..........

Black6      GGAGCTTAGA GAAATCTATA AAACTAGGTC CATATACTGC CCTAAAGAGC TGGGAGAACC ATGACTTGGC
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      TCCCATTCTT CTGCATTCAC CAGCTGACTG TGCCCAAGAC ---------- ---------- ----------
domesticus  .......... .......... .......... .......... ---------- ---------- ----------
musculus    .......... .......... .......... .......... TGCCGCAGAC CTTTTGGGTC CCTATGTCTG

Black6      ---------- ---------- ---------- ---------- ---------- ---------- ----------
domesticus  ---------- ---------- ---------- ---------- ---------- ---------- ----------
musculus    CGTGGAACGG GGTCTCTCGA CGCGGGTGGG CAAAGCGTGG ATGATGACAG ACAGACACAC ACGCAGGAGA

Black6      ---------- ---------- ---------- ---------- ---------- ---------- ----------
domesticus  ---------- ---------- ---------- ---------- ---------- ---------- ----------
musculus    GTTTGTGTGG AATCTGAATG TAATTTTACA ACTGAGCATC AGACTTTTTA TGCAGAGGAC AATAAGGAAG

Black6      ---------- ---------- ---------- ---------- ---------- ---------- ----------
domesticus  ---------- ---------- ---------- ---------- ---------- ---------- ----------
musculus    TTGGGTGACA TATTCGCAAG GTACAATTGA GGTAACTGGA ATCTTACATA AAACAGAGGA ATGCAAACAC

Black6      ---------- ---------- ---------- ---------- ---------- ---------- ----------
domesticus  ---------- ---------- ---------- ---------- ---------- ---------- ----------
musculus    AAAAGGTCTA ACGGGAACCA CCCAGGATAG AATACATTGA TAACAGCTGG GATCAGCAAG GGCTCCACCT

Black6      ---------- ---------- ---------- ---------- ---------- ---------- ----------
domesticus  ---------- ---------- ---------- ---------- ---------- ---------- ----------
musculus    AAAATTTACT AATCTTAGAA GCCAGGGTCA AGGGCTTCAT GTCCTTGCCA TAGTTCCTAT TTTAGTCTCT

Black6      ---------- ---------- ---------- ---------- ---------- ---------- ----------
domesticus  ---------- ---------- ---------- ---------- ---------- ---------- ----------
musculus    TGTATAGTCC ACCTTCCCCT TAGGCCATTG TAAATACGTG TGTATGGGTG TAACTCTGCT ATAGTATCTA

Black6      ---------- ---------- ---------- ---------- ---------- ---------- ----------
domesticus  ---------- ---------- ---------- ---------- ---------- ---------- ----------
musculus    TTCTGGTTTC TCCTTTGGTC TGAAACTTCC TTCTTCCTAA GTGATGGTAA ATTCCTGTGT AAGGGAGTGA

Black6      ---------- ---------- ---------- ---------- ---------- ---------- ----------
domesticus  ---------- ---------- ---------- ---------- ---------- ---------- ----------
musculus    CTTAGCTTTT ACAATCTTGC TGAATTCTAA ACCCTTGATC TTGGTCAAGG ATGCCCTTGG ACTGTTGGAA

Black6      ---------- ---------- ---------- ---------- ---------- ---------- ----------
domesticus  ---------- ---------- ---------- ---------- ---------- ---------- ----------
musculus    CACTGGCGGA AGGCTTTAGC CATGTCAGAG CACTTATAAT AGGAAAATAC TGAAAGAGAG CACGTGGATC

Black6      ---------- ---------- ---------- ---------- ---------- ---------- ----------
domesticus  ---------- ---------- ---------- ---------- ---------- ---------- ----------
musculus    CATACTCTAG ACTAGTGCGG TAATGGAACT TGGGAATGAA AAATTAATAT ATGGGTAGAG AACACTAGCC

Black6      ---------- ---------- ---------- ---------- ---------- ---------- ----------
domesticus  ---------- ---------- ---------- ---------- ---------- ---------- ----------
musculus    TCCAGGAGGA GAGCTTCCTT GAAACTCTTT TGCCTCATGA GTGACTTTTA AGCTTTTCGG CTTGTCCAGT

Black6      ---------- ---------- ---------C TTATGGATCT ATGAGAAATT AGTCCCCACT GAGGCCTAAG
domesticus  ---------- ---------- ---------. .......... .......... .......... ..........
musculus    TGACTCGACC AGAGAGCGTA GCACAAGAC. .......... .......... ...Y...... ..........

Black6      CCTATTGACA GACTAAGATG ACCAAGTCTA ATGTGTGAGG TCCAGGCAAG TAAGGTCCAA TTGTAACCAG
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........
```

```
Black6      TGAGTCTGTC CCAAGACAGC ACGGTTCCCT CCAAACTAAG AAGAAGGGTA GCATGTCAGG AGCCATAATG
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      GGACAAGCTA ATAACTAGCA GGTGATCATC CTGAAGTGGC TTGCTTCAGA TTATTTTATA ACCTGTGACA
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... ..R.....Y. ..........

Black6      GATGAGCCCT TATTAAGCAA GGCTGTAAGG GACTCATTTT AGGAAATATT TCTGCTATTA ATATGCTTTT
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    ......M... .......... .......... .......... .......... .......... ..........

Black6      CCTCTTATTA AACATATATC ACTAAGTAGT AGCACACTCC TTTGGAGCAA ATCTCTGCAG ATCCATGAAG
domesticus  .......... .......... .......... .......... .........R .......... ..........
musculus    .......... .......... .......... .......... .........G .......... ..........

Black6      ATGTACAATC TTGTAGTGAT GCTATATAGA CAAATAGATA ACTTAAGTCT TAATGATGAT CCTATAAGAA
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... ....XX.... .......... ..........

Black6      TTCCTAAAAT TATATCTGTC ATTACTAAGC TCTTTTATAG TGGGACTTTT TTTCTTTTAA TTTTTTTA-T
domesticus  .......... .......... .......... .......... .......... .......... ........-.
musculus    ........R. .......... .......... .......... .......... .......... .....XXXA.

Black6      TAGGTATTTT CCTCATTTAC ATTTCTAATG CTATCCCAAA AGTCCTCCAT ACCCTCCCCC CAACTCTCCT
domesticus  .......... .......... .......... .......... .......... .......... -.........
musculus    ....Y..... .......... ...Y...... .......... .....Y.... .......... ......Y...

Black6      ACCCACCCAC TCCCACTTCT TGACCCTGGC ATTCCCCTGT ACTGAGGCAT ATAAAGTTTG CAAGACCAAT
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... R......... .......... .......... ......S...

Black6      GGGCCTCTAT TTTCACTGAT GGCTGACTAT GCCATCTTCT GATACATATG CAGCTAGAGA CACAAGCTCC
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .R........ .......... .........K .......... .......... .......... ..........

Black6      CGGGGGGGGG GGGGGGTACT AGTTAGTTCA TATTGTTGTT GTTCCACTTA TAGGGTTGCA GATCCCTTTA
domesticus  .......... ...--..... .......... .......... .......... .........Y. ..........
musculus    G...A.---- ------.... RR........ .......... .......... .......... ..........

Black6      GCTCCTTGGG TACTTTCTCT AGCTCCTCCA TTGGGGGTCC TGTGATCCAT CCAATAGCTG ACTGTGAGCA
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......Y.. .......... .......... ..........

Black6      TCCACTTCTG TGTTTGCTAG GCCCCGGCAT ATAGTCAAAC CTGCAATGAG AACTCTGCCA GTCTCCCTCG
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ........Y.

Black6      TGTCACCAGT TAATTGCTCT TAGATGGTAA CCAGACTTTC TCCTACACAG AGCATATTCC AAGAGTTATA
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    ..W....... .......... .......... .......... .......... .......... ..........

Black6      TAACAATTAA CCAAAGGTCA TTAAAAGGGA ACTAATGATT TCTTATAGGT GCTAGGACAG AAGATAAAAT
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .....K.... .......... .......... .......... .......... ..........

Black6      ATTGACTGTG TTTATCTATA TAAAACTTCA CTTATAACTT AGTTATGGTT TCAAACCTTT GGTCAACCCG
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .Y........ ........Y.

Black6      TGGAGCTGAG ACAGATGATG GATGCTTAGC TAGATAATTA CTCGTAATGA ATTTGCATGT AAACATTCTC
domesticus  .......... .......... .......... .......... .......W.. .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      TGTTGTAAGC TTCTATTTCA ATTTATGATT TGATTTTCTA TGTGAAATCT TGATGAACTT TTCAACATGT
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... K......... .......... .......... ........Y. .......... ..........
```

```
Black6      GATCATGTAT TCTGAATGAT GTTTAAGAGT GGAGGACAAA GAGAAGAGTC AGAACTGAGG GACTGAGGTG
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ......R...

Black6      AGAGGAATGG AAGTGAGAAC TGAGCTGAGA GAGGAACTGA GCTGAGGAGT TTTTAGACAG AACACTTTAG
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      ATTAGAAAGA GAATGCAGAG TGGAGTATCT TAGAAATTAT AGGTAACATA AAATACTAAG AGAGCAATGT
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... ....R..... .......... .......... .......... ..........

Black6      GCAGAATGCA GAAGGAAAGG GGAAAAAAGA TGTTGCAGAG AGCAGAGGGA GCAGGCAGGC TTTTCCTGAC
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      CATGTTTTAG TCTTATTAAA AGAAACAAAG CTTTTTCCCT ACAAACATGG GTTTAATTCA TTCAGCATTA
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    ...R...... .S........ .......... .......... .......... .......... ..........

Black6      AAAAGGTAGA AATCTTTTCT TTTTCTTTGT AATAAAGATT AGAGTGCAGT TTTCATCCAG AATAAGTGAG
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... ..Y....... .......... .......... .......... .......... ..........

Black6      TTCGTTCTGT ACTGTGCTTG GTCTTTTTCT CCATATGCAT ATGTTAGTAT GGATGTGTGT GTAGCTGGGC
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... ..M....... ..........

Black6      CCAGTTGGGT TTGAATGGAA ATGTATAAAT GAATAAATGA GCATGTATGA ATCTGTATAT GAACTTATGT
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      GTGATTATGC ATATTTGCTT ATGTCAAAGT TTTTCCTTCT GCGAATGTTA TGAGTGTTAT TCTCTTCTCC
domesticus  .......... .......... .......... .......... ......R... .......... ..........
musculus    .......... .......... .......... .......... R......... .......... ..........

Black6      TGGTTCAATA GAAGTTTATT GCTCCAAAGT CCCCCCTAGT CTGACAAGCA AGAGGCATCT GGACAATAGG
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... ......M... ..........

Black6      GAAAAGTCTC TAGCAATTAA TCCTCTACTT ATTCTCCTGC TGTTTTAGGG TGAGGTGCTA AGTGCAGTAA
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      GTGCAGTTTC TGCCCTCAGA AGATCACAGA AGGCAGGTCA GCTACAATAG CATAGTAACT TACATAAGTA
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... ..K....Y.. .Y..S..... .......... ..........

Black6      AACTTTTAAT TATACAGCAA CCCTAAATTT CTATTAAGAC AAAGTCTTAC CACAGAATCT CCCAAGACAA
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      AGTCTTACTC TCTGCCTACA CTCTTCCACC TCATCTGCCT GAAGAAGAAT GGATAAGAAA GAAGAAACAT
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........

Black6      CCCCTTTGGG GCTGGCCTTC AGCGTCAGGT AGCTCTTCTT GCTAAACAAT ACAGGAAGAA ATCTTCTGGG
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........
Black6      GATCCCCACA TTTGCATAAA AGACACTGAG TGTGCTCTCA AAGGGCCACC CCCTCACTAG AGAGATAAGG
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... ...R...... .......... .......... .......... .......... ..........

Black6      TGCTCTGTGT TTCATAATTG TTATTCCTGA GATTTTAATC AGTGTAGACT CCTACACAGG AGAAAGTACC
domesticus  .......... .......... .......... .......... .......... .......... ..........
musculus    .......... .......... .......... .......... .......... .......... ..........
```

```
Black6      TGGCACTATA TAAGGCACTG AGCTCAAGTC CCTCTGCATC TCTGCACCTC ACCAGGCATC AGTCATGAAG
domesticus  .......... .......... ---------- ---------- ---------- ---------- ----------
musculus    .......... .......... ---------- ---------- ---------- ---------- ----------

Black6      ATCCATTACC TGCTCTTTGC CTTTATCCTG GTGATGCTGT CTCCACTTGC AG
domesticus  ---------- ---------- ---------- ---------- ---------- --
musculus    ---------- ---------- ---------- ---------- ---------- --
```

13 Digital supplement

- RACE sequences, raw data
- Agilent Bioanalyzer data of Testis RNA analysis (4 LP10 knock-out mice, 4 wild-type mice Black6)
- Affymetriy GeneChip Mouse Genome 430 2.0 raw data (4 LP10 knock-out mice, 4 wild-type mice Black6)
- LP10 targeting vector sequence
- Luciflip plasmid sequence

14 Danksagung

An dieser Stelle möchte ich meinen Dank aussprechen an:

Diethard Tautz, der mir die Gelegenheit gab an diesen interessanten Projekten zu arbeiten.

Peter Nürnberg, der sich bereiterklärt hat das Zweitgutachten zu übernehmen, und mich insbesondere bei den Microarrayexperimenten unterstützt hat.

Hartmut Arnd, der sich bereit erklärt hat den Vorsitz meiner mündlichen Prüfung zu übernehmen.

Wim Damen, für die freundschaftliche Unterstützung die ich durch Ihn während meiner Doktorarbeit erfahren habe.

Daniela Häming, die eine riesengroße Unterstützung bei dieser Arbeit war.

Fabian Staubach und Bettina Harr, die die beiden Kandidatengene LP10 und Mkk7 als solche entdeckt haben und somit die Grundlage für diese Arbeit geschaffen haben. und die immer hilfsbereit und offen für Diskussion waren.

Alistair McGregor für das Korrekturlesen des englischen Textes.

Eva Siegmund, Petra Neumann und Gerti Meyer zu Altenschilde für die große Hilfe in administrativen Angelegenheiten.

Jonathan Howard und Julia Hunn, die mir ihr Zellkulturlabor und das entsprechende Setup zur Verfügung gestellt haben.

Pia Scholl, Julia Pfeiffer, und Tanja Tropartz die mir bei der Histologie geholfen haben.

Die Mitglieder der AGs Tautz und Damen für die gute und freundschaftliche Zusammenarbeit.

15 Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit − einschließlich Tabellen, Karten und Abbildungen −, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Herrn Professor Dr. Diethard Tautz betreut worden.

Köln, den 13.10.2008

Tobias Heinen