

Regulatory elements controlling
the expression of OR37 genes

**Dissertation zur Erlangung des Doktorgrades
der Naturwissenschaften (Dr. rer. nat.)**

**Fakultät Naturwissenschaften
Universität Hohenheim**

Institut für Physiologie

vorgelegt von

Yongquan Zhang

aus Zhanglou Village – China

2007

Erklärung

Ich versichere, dass ich diese Dissertation selbständig gemäß der Promotionsordnung angefertigt, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche kenntlich gemacht habe.

Yongquan Zhang

CONTENTS

1.	Introduction	1
2.	Materials and methods.....	4
2.1	Materials.....	4
2.1.1	Animals.....	4
2.1.2	Reagents.....	4
2.1.3	Enzymes and vectors.....	4
2.1.4	Kits for molecular biological techniques.....	4
2.2	Methods.....	5
2.2.1	General methods for preparation, purification, recombination and characterization of DNA material.....	5
2.2.1.1	Preparation of genomic DNA.....	5
2.2.1.2	Preparation of plasmids.....	5
2.2.1.3	Restrict digestion of DNA.....	6
2.2.1.4	Ligation.....	6
2.2.1.5	Transformation.....	6
2.2.1.6	Agarose gel electrophoresis.....	6
2.2.1.7	Purification of DNA from agarose gel.....	7
2.2.1.8	Sequencing.....	7
2.2. 2	Construction of plasmids and transfection.....	8
2.2.2.1	Promoter subcloning.....	8
2.2.2.2	Subcloning of mOR37C promoter and site directed mutagenesis of O/E binding sites.....	8
2.2.2.3	Single copy of mOR120-1 promoter subcloning and site directed mutagenesis of LHX-2 binding site.....	9
2.2.2.4	H-enhancer subcloning.....	9
2.2.2.5	Transcription factors subcloning.....	10
2.2.2.6	Cell culture and transfection.....	11
2.2.2.7	Measurement of luciferase activity.....	11
2.2.3	Bioinformatic studies.....	11

2.2.3.1	Download of genomic sequence.....	11
2.2.3.2	Formatting of genomic sequence.....	11
2.2.3.3	Comparative analysis of genomic sequence by PipMaker.....	12
2.2.3.4	Search for transcription factor binding sites.....	12
2.2.4	MOR37C transgenic mouse.....	12
2.2.4.1	Plasmid construction.....	12
2.2.4.2	Generation of transgenic mice.....	12
2.2.5	PCR based methods.....	13
2.2.5.1	Real-time PCR.....	13
2.2.5.2	Chromosome conformation capturing (3C).....	13
2.2.5.3	DNA walking.....	14
2.2.5.4	Inverse PCR.....	15
2.2.5.5	Adaptor mediated PCR.....	16
2.2.6	Fluorescence in situ hybridization (FISH) on metaphase chromosomes.....	17
2.2.7	X-gal staining.....	18
2.2.8	Immunohistochemistry.....	18
2.2.9	Cell counts.....	19
2.2.10	Microscopy and photography.....	19
3.	Results.....	20
3.1	Functional interaction between mOR37 promoter and transcription factors in a heterologous system.....	20
3.1.1	Background.....	20
3.1.2	The strategy for cotransfection.....	20
3.1.3	Monitoring the interaction of Lhx-2 and putative mOR37 promoters by luciferase expression.....	22
3.1.4	Site-directed mutagenesis of bases flanking the Homeodomain-like site in the promoter mOR120-1.....	23
3.1.5	Functional interaction of O/E-factors with different promoters.....	24
3.1.6	Site directed mutagenesis of the olf-1 site in the mOR37C promoter	29
3.1.7	Simultaneous activation of the mOR37B promoter by O/E-2 and Lhx-2.....	31

3.1.8	Effect of the H element on O/E-2 interaction with mOR37B and mOR120-1 promoters.....	32
3.2	<i>In vivo</i> demonstration of the role of the promoter in regulating the topographic expression of mOR37 gene.....	34
3.2.1	Expression of mOR37C transgene.....	34
3.2.2	Tissue specificity of transgene expression.....	40
3.2.3	Slight delay of the onset of transgene expression.....	40
3.2.4	Integration site and copy number of transgenes.....	41
3.2.5	Mutually exclusive and monoallelic expression of the mOR37C transgene.....	44
3.2.6	Projection of transgene expressing OSNs to the olfactory bulb.....	45
3.2.7	Effect of an ectopic mOR37C expression on the projection.....	49
3.2.8	Search for the integration site of the transgene in line7.....	52
3.3	No detection of interaction between H element and mOR37 promoter by Chromosome Conformation Capture (3C).....	55
3.3.1	The principle of 3C.....	55
3.3.2	Searching for an interaction between H element and mOR37C promoter.....	56
3.4	Search for locus control region-like elements of the mOR37 cluster	60
3.4.1	Background.....	60
3.4.2	Unavailability of identifying mOR37 cluster I related LCR like elements by sequence comparison across closely related species	60
3.4.2.1	Conservation of mOR28 cluster across multi-species.....	63
3.4.2.2	Does sequence comparison allow to identify H element sequences in ancient species?.....	64
3.4.2.3	Is the “H element” in dog related to OR28 cluster or TCR gene cluster?.....	66
3.4.2.4	Attempts to identify an H element sequence by means of multi sequence comparison.....	66
3.4.3	Comparison of mOR37 gene clusters from different species.....	69
3.4.4	<i>Pipmaker</i> analysis of the cluster I locus plus opossum.....	72
3.4.5	<i>Pipmaker</i> analyses of the cluster II locus.....	75
3.4.6	Comparison of the Cluster I and Cluster II locus.....	78
3.4.7	Comparison of the conserved OR37 segment and the H element...	84

4.	Discussion.....	87
5.	Summary.....	97
6.	Zusammenfassung.....	99
7.	References.....	101
8.	Abbreviations.....	112

1. Introduction

The capability of the mammalian olfactory system to detect a vast array of small volatile compounds is mediated by more than 1000 different isoforms of G-protein coupled odorant receptors (ORs) (Buck and Axel, 1991) which are encoded by the largest gene family in vertebrate genomes (Zhang and Firestein, 2002; Young and Trask, 2002; Mombaerts, 2004; Godfrey et al., 2004; Malnic et al., 2004). Each of the several million olfactory sensory neurons (OSNs) in the nasal neuroepithelium expresses just a single gene from this large repertoire, which renders them selectively responsive to distinct chemical compounds (Malnic et al., 1999; Touhara et al., 1999; Bozza et al., 2002). From the two alleles that code for each receptor, only one is selected per cell (Chess et al., 1994; Ishii et al., 2001); the reason for this monoallelic expression of OR genes is not yet fully understood. Interestingly, OSNs which express a given OR are not randomly dispersed throughout the olfactory epithelium (OE) but restricted to a defined zone. By determining the patterns of a small set of OR genes the epithelium has originally been divided into three or four separate zones (Vassar et al., 1993; Ressler et al., 1993; Strotmann et al., 1994b); more recent data, however, have provided evidence that each OR gene might have a distinctive, subtype-specific zonal pattern (Iwema et al., 2004; Miyamichi et al., 2005). Remarkably, dependent on the OR they express and their position in the OE, OSNs send their axon to one of two target glomeruli in the olfactory bulb (OB), one positioned in the medial hemisphere in the bulb, the other in the lateral hemisphere (Vassar et al., 1994; Ressler et al., 1994; Mombaerts et al., 1996; Wang et al., 1998; Levai et al., 2003; Feinstein and Mombaerts, 2004); for recent reviews, see Mombaerts (2006) and Strotmann and Breer (2006).

The regulatory DNA sequences that are required for the singular expression of ORs in a topographical manner are still largely elusive. Previous observations that almost all genes coding for ORs are organized in clusters, rather than being homogeneously dispersed throughout the genome have led to the idea that some aspects of receptor gene expression might derive from transcriptional control at the level of the OR gene cluster. Genes which share the same expression pattern in fact tend to be linked together at the same locus (Malnic et al., 1999; Zhang and Firestein, 2002; Miyamichi et al., 2005). Furthermore, approaches using transgenic mice have demonstrated that a DNA element located in neighborhood to an OR gene cluster participated in the expression control of the corresponding OR genes

(Serizawa et al., 2000; Serizawa et al., 2003); due to its homology in mouse and human it was named 'H-region'. Very recent experiments provided evidence that this DNA-element might act globally to activate OR gene expression (Lomvardas et al., 2006). These results supported the concept of a locus dependent regulation of OR gene expression, probably involving a 'locus control region' (LCR) similar to the one found e.g. close to the beta-globin gene cluster (Grosveld et al., 1987; Forrester et al., 1987). In contrast to these data, however, other OR genes apparently require only a rather short genomic region surrounding their coding sequence to drive expression in a tissue and cell specific manner (Qasba and Reed, 1998; Vassalli et al., 2002; Rothman et al., 2005), favoring the concept that regulatory elements located immediately upstream of the transcription start site (TSS) of each gene are sufficient for a correct expression, independent of the cluster context.

Some transcription factors have been reported to be involved in regulation of OR expression. One group of novel helix-loop-helix (bHLH) proteins, notably O/E-1, O/E-2, O/E-3 and O/E-4 were reported to be expressed in immature and mature olfactory sensory neurons and are considered to regulate the neuronal development by controlling the expression of specific genes including ORs (Wang et al., 1993; Liberg et al., 2002; Wang et al., 1997; Wang et al., 2002; Wang et al., 2004). Another candidate which seems to participate in the development of OSNs is the LIM-homeoprotein Lhx-2, whose deficiency blocks the differentiation of OSN and causes the silence of some olfactory receptor genes (Hirota and Mombaerts, 2004). For both O/Es and Lhx-2 mutations in their binding sites within the M71 OR gene promoter reduced the number of OSNs expressing M71, implying a direct involvement of O/Es and Lhx-2 in controlling OR gene expression (Rothman et al., 2005).

The mOR37 genes belong to family *mOR262* (according to the nomenclature of Zhang and Firestein 2002) and represent a small group with 12 members in the mouse genome (Strotmann et al., 1999; Hoppe et al., 2003a; Hoppe et al., 2003b). Distinct from the other ORs all mOR37 proteins possess a unique insertion of six amino acids in the third extracellular loop of their seven transmembrane construct (Kubick et al., 1997). They exhibit an unusual spatial expression pattern in the OE, being expressed exclusively in OSNs that are restricted to a small patch in the center of the turbinate (Strotmann et al., 1992; Strotmann et al., 1994a; Kubick et al., 1997). This pattern is quite different from the standard zonal expression rule. As another unique feature, OSN populations expressing an mOR37 gene do not send their

axons to two glomeruli in the OB, but only to a single glomerulus (Strotmann et al., 2000). The molecular bases for these characteristics are not known. The observation that all genes encoding the mOR37 receptors are linked together at two clusters on chromosome 4 and each cluster contains only members of this receptor family (Hoppe et al., 2000; Hoppe et al., 2003b) led to the idea that their affiliation to these loci is required for the unique expression pattern. On the other hand, each individual mOR37 gene possesses a DNA element (putative promoter) immediately upstream of its TSS which is characteristic and unique for those genes expressed in the patch (Hoppe et al., 2003b; Hoppe et al., 2006), opening the possibility that each one is controlled autonomously, independent from that particular cluster context.

To obtain further insight into the principles and mechanisms underlying the topographically restricted expression of the mOR37 genes, three strategies have been employed: cotransfection of the putative mOR37 gene promoter and transcription factors into HEK293 cells to examine the functional interaction *in vitro*; generation of transgenic mice to investigate the function of the putative promoter of mOR37C (mOR262-12; Olfr157) *in vivo*; and comparative analysis approach to search for LCR (Locus Control Region)-like elements for the OR37 gene cluster.

2. Materials and methods

2.1 Materials

2.1.1 Animals

Wild type C57/J6 mice were purchased from Charles River (Sulzfeld, Germany). All the mouse treatment complies with Principles of Animal Care, publication no.85-23, revised 1985, of the National Institutes of Health and with the current laws of Germany.

2.1.2 Reagents

5-bromo-4-chloro-3-indolyl- β -D-galactodise (X-gal) and Isopropyl- β -D-thiogalactopyranoside (IPTG) were purchased from Biomol. LB (Luria Bertani)-broth was purchased from International Diagnostic Group (IGD) Agarose was obtained from Invitrogen Inc. HEPES (4-2-hydroxyethyl-1-piperazineethanesulfonic acid) and TRIS (tris(hydroxymethyl)methylamine)-HCl were purchased from Roth. EDTA (ethylenediaminetetraacetic acid), glycerol, Tween 20, paraformaldehyde, DTT (dithiothreitol), EtBr (ethidium bromide) and DMSO (dimethyl sulphoxide) were purchased from Sigma.

2.1.3 Enzymes and vectors

T4 ligase, restriction enzymes SpeI, NotI, PmeI, PaeI, EcoRI, SacI, XbaI, HindIII, KpnI, ScaI, Sall, BamHI, NcoI, XhoI, NspI, aflII and NsiI were purchased from New England BioLabs. PGEM-T vector was obtained from Promega. PGL3-basic and pcDNA3.1 vectors were purchased from Invitrogen. pBluescript II KS(+) was obtained from Fermentas.

2.1.4 Kits for molecular biological techniques

PCR kit with Taq Polymerase was purchased from Q-Biogene. PCR kit with PWO Polymerase was obtained from PeQlab. Long Range PCR kit was purchased from QIAGEN. Site directed mutation kit 'Pfu Turbo' was purchased from Stratagene. DIG Labelling PCR kit and DIG Random labelling kit were purchased from Roche. Plasmid Mini and Midi preparation Flexi Pre Kits were obtained from QIAGEN. DNA purification kit GENE CLEAN^{RII} was obtained from Q-Biogene and Perfectprep Gel Cleanup was purchased from Eppendorf. Cell Culture transfection kit LipofectaminTM2000 was obtained from Invitrogen and Luciferase Assay System was purchased from Promega. DNA Walking kit was purchased from Seegene.

2.2 Methods

2.2.1 General methods for preparation, purification, recombination and characterization of DNA material

2.2.1.1 Preparation of genomic DNA

1 cm of mouse tail was immersed in 500ul of Lysis buffer (0,1M Tris-HCl pH8,0, 5mM EDTA, 0,2% SDS, 200mM NaCl) and incubated overnight at 55°C with 30µl of 10mg/ml Proteinase K. After repeated inversion and cooling down to room temperature undigested material was removed by centrifugation at 13,000 rpm (Eppendorf 5417R) for 8 minutes at 4°C. The supernatant was transferred to a new tube and 250ul of saturated NaCl (6M) was added to precipitate proteins for 5 minutes at room temperature. The proteins were spun down at 13,000 rpm for 10 minutes at 4°C and the supernatant was mixed with 500µl of isopropanol by inversion in a new tube till the white thread materials appeared. Following precipitation by centrifugation at 13,000 rpm (Eppendorf 5417R) for 10 minutes at room temperature the genomic DNA was washed with 500µl of 70% ethanol and air dried for 50 minutes at room temperature. The pellet was dissolved in 50µl molecular grade water containing 200ug/ml RNase A overnight at 4°C.

2.2.1.2 Preparation of plasmids

Plasmids were prepared using the Flexi Pre Kit. 1,5ml of overnight culture from a single bacterial colony was centrifuged at maximal speed (14,000 rpm, Eppendorf 5417R) for 30 seconds to pellet the cells. Following resuspension of the cells in 200µl of solution I (100mM Tris-HCl pH7.5, 10 mM EDTA, 400ug/ml RNase I) by vigorous vortexing 200µl of solution II (1M NaOH, 5.3 % (w/v) SDS) was added to lyse the cells. After incubation for 5 minutes at room temperature 200µl of solution III (3M potassium acetate) was added and mixed by inverting the tubes for several times, followed by centrifugation at maximal speed at room temperature for 5 minutes. The supernatant was transferred to a new tube and mixed with 0.7 volume of isopropanol. After incubation for 10 minutes at room temperature the plamid was precipitated by centrifugation at maximal speed for 10 minutes. 100ul of sephaglas FP slurry (in buffered solution of 7M Guanidine-HCl, 50mM Tris HCl PH7.5, 10 mM EDTA) was added to the pellet and vortexed for 1 minute. The mixture was centrifuged at maximal speed for 15 seconds and the pellet was first washed with 200ul of Wash

buffer (20mM Tris-HCl pH8,0, 1mM EDTA, 0,1M NaCl, to which 18ml of absolute ethanol was added) and then with 300µl of 70% ethanol. The sparse pellet was air dried till becoming white and 50ul of sterile water was added to resuspend the pellet. After incubation for 5 minutes at room temperature the suspension was centrifuged at maximal speed for 5 minutes and the supernatant was transferred to a new tube.

2.2.1.3 Restrict digestion of DNA

For plasmid and PCR product 3units/ug was used and the incubation lasts for no longer than 1 hour. For Genomic DNA 6units/ug was employed and the digestion was performed at least for 4 hours, normally overnight. The digestion was performed according to the manufacture's instructions.

2.2.1.4 Ligation

The ligation was performed with T4 ligase in 1x ligation buffer (30mM Tris HCl pH7,8, 10 mM DDT, 1mM ATP and 5% PEG) overnight at 16°C. The molecular ratio of the cloning vector and the digested fragment was adjusted around 1:3.

2.2.1.5 Transformation

3 Single colonies of E.coli XL 1 blue were inoculated to 50 ml of LB medium and incubated overnight with a G24 environmental incubator shaker (New Brunswick Scientific.co.in). After inoculating 5 ml of overnight culture to 500 ml of new LB medium in a barrier flask and growing at 37°C till the optical density (O.D) was around 0,5 the culture was centrifuged for 15 minutes at 5,500g in Beckmann coulter centrifuge. The pellet was washed two times in ice cold sterile waster and dissolved in 2 ml of ice cold 10% glycerol in HEPES, followed by separating the cells into 100 µl of cell aliquots and storing at -70°C. After the electrocompetent cells were defrozen on ice for 15 minutes and mixed with 1/3 volume of the ligation solution for 2 minutes the cells were transferred to a pre-chilled 0,2cm cuvette and the transformation was performed using a Gene Pulser™ (BIO-RAD) set at 25 uF and 2,4 kV. The electroporated cells were incubated for 1 hour at 37°C and spread on LB plates supplemented with proper antibiotic.

2.2.1.6 Agarose gel electrophoresis

Genomic DNA, plasmid or PCR products were electrophorised on 0,8% to 1% agarose gel containing ethidium bromide for visualization of DNA under ultraviolet light (302nm). 10 µl of DNA solution was mixed with 2 ul of 6x loading buffer (0,0025% bromophenol blue, 0,025% xylencyanol and 30% glycerol) and

electrophorised at 100-150 V using 1x TAE buffer (90mM Tris-HCl pH7,4, 2mM EDTA pH8,0 and 90 mM acetic acid). An appropriate DNA molecular weight maker (100bp or 1Kb) was loaded to determine the size of the DNA materials and the gel was subject to photography using a gel documentation system (Labortechnik).

2.2.1.7 Purification of DNA from agarose gel

Purification of DNA fragments smaller than 3Kb was performed using the GENE CLEAN[®]II kit. The excised band from an agarose gel was melted with 3 volumes of NaI solution for about 10 minutes at 50°C. 6µl of glass milk suspension was added to the solution, vortexed and incubated for 5 minutes at room temperature with frequent rocking, followed by centrifugation at maximal speed for 5 seconds. The resulting pellet was washed three times with New Wash and resuspended in 20-30µl of sterile water. After centrifugation at maximal speed for 6 minutes the eluted DNA was transferred to a new tube and then subject to examination by agarose gel electrophoresis. Perfectprep Gel Cleanup kit was used to purify fragments larger than 3 Kb from agarose gel. The gel slice containing the DNA band was melted with 3 volumes of Binding buffer by incubating for 10 minutes at 50°C. 1x original gel slice volume of isopropanol was added to the solution and mixed by inversion, followed by centrifugation through the Spin Column at 7,000g for 1 minute. 750µl of wash buffer was added to the column and centrifuged at 7,000g for 1 minute. After an additional minute of centrifugation at 7,000g the DNA was eluted in 30 µl of Molecular Biology Grade Water by centrifuging at 7,000g for 1 minute.

2.2.1.8 Sequencing

The ABI BIO automatic sequencer 310 and ABI Big-Dye-Terminator cycle sequencing kit (Kerkin Elmer, Applied Biosystems) was used to perform DNA sequencing. The sequencing reaction was run in 10 µl of solution containing 400ng plasmid or 20-50 ng of purified PCR product (for direct sequencing), 1µl of 5µM primer, 1µl of 5x sequencing buffer and 2 µl of Big-Dye with the program: 1 cycle of 30 seconds at 96°C, followed by 24 cycles of 10 seconds at 96°C, 5 seconds at 50°C and 4 minutes at 60°C. The DNA was precipitated by adding 2,5 volume of 100% ethanol, 0,1 volume of 3M sodium acetate (pH5,2) and 125mM EDTA (pH8,0) and then pelleted by centrifugation for 30 minutes at 3,000g at 4°C. The pellet was washed with 70% ethanol, dissolved in 25 ul of Licosolv and subjected to sequencing. The sequence document was analyzed with Sequencer 4.1.1 or Chromas 1.41 program or by BLAST.

2.2. 2 Construction of plasmids and transfection

2.2.2.1 Promoter subcloning

PGL3-basic plasmids were obtained by transforming JM109 competent cells with the commercial vector and subsequent Midi preparation. The KpnI/HindIII linker was generated by annealing the sense oligo strand and the antisense strand in annealing buffer (100mM KCl, 10mM Tris PH8,0, 1mM EDTA) for 1 minute at 95°C, 30 minutes at 65°C , 8 minutes at 37°C and 10 minutes at room temperature. The annealed linker was then ligated with KpnI/HindIII digested PGL3-basic plasmid and the resulting PGL3-basic/Linker was digested by EcoRI/SacI or EcoRI/XhoI. The HISi-I vector containing the mOR37A promoter region or mOR120-1 promoter region was kindly provided by Mr. Henning Frank (Institute of Physiology, University of Hohenheim) and pBluescript II KS(+) vector containing mOR37B promoter region was kindly supplied by Dr. Reiner Hoppe. The EcoRI/SacI restricted mOR37A, mOR120-1, mOR11-1 and mOR18-2 promote fragments or EcoRI/XhoI restricted mOR37B promoter fragment was accordingly subcloned into the PGL3-basic vector immediately upstream of the cDNA of Luciferase gene. The sequences of the oligonucleotide strands for the linker are as follows:

Sense strand: 5'-cga att cgg cac tag tga gct cct cga ga-3'

Antisense strand: 5'-agc ttc tcg agg agc tca cta gtg ccg aat tcg gta c-3'

2.2.2.2 Subcloning of mOR37C promoter and site directed mutagenesis of O/E binding sites

SpeI/KpnI fragment from mOR37C transgenic vector (see 2.2.4.1) was purified by GENECLAN^{RII}, ligated with SpeI/KpnI restricted PGL3-basic-linker, and subsequently electrotransformed into JM109 competent. The PCR for mutation was then carried out in 50µl of reaction solution containing 50ng of PGL3 vector containing one copy of mOR37C promoter as template, 125ng of Forward primer and Reverse primer respectively, 200uM dNTP, 5µl of 10X reaction buffer and 1µl of Pfu polymerase and processed with one cycle of 30 seconds at 95°C, 18 cycles of 30 seconds at 95°C, 1 minute at 55°C, 6 minutes at 68°C, 1 cycle of 4 minutes at 68°C. After digestion with 1ul of DpnI for 1 hour at 37°C to remove the template the PCR products were transformed into JM109 competent cells. The mutant plasmids were

finally identified by sequencing with PGL3-basic primer. The sequences of the primers for site-directed mutation are as follows:

Forward primer: 5'-cct tgt gat caa atg aaa gga gaa ttc tga g-3'

Reverse primer: 5'-ctc aga att ctc ctt tca ttt gat cac aag g-3'

2.2.2.3 Single copy of mOR120-1 promoter subcloning and site directed mutagenesis of LHX-2 binding site

One copy of the mOR120-1 promoter was removed from the PGL3-promoter vector by HindIII digestion and the vector containing one copy of mOR120-1 promoter was purified and religated with T4 ligase. The PCR for mutation was then carried out in 50µl of reaction solution containing 40ng of PGL3 vector containing one copy of mOR120-1 promoter as template, 125ng of Forward primer and Reverse primer respectively, 200µM dNTP, 5 µl of 10X reaction buffer and 1µl of Pfu polymerase and processed with one cycle of 30 seconds at 95°C, 12 cycles of 30 seconds at 95°C, 1 minute at 55°C, 5,5 minutes at 68°C, 1 cycle of 4 minutes at 68°C. After digested with 1µl of DpnI for 1 hour at 37°C to remove the templates the PCR products were transformed into JM109 competent cells. The correct colonies were identified by EcoRI/HindIII digestion which generates one copy of mOR120-1 promoter and further digestion by PacI which linearizes the nonmutant vector but has no restriction function on mutant vector. The sequences of the primers for site-directed mutation are as follows:

Forward primer: 5'-cct ctc aga act tgg gtt gat taa gaa cac ata cct c-3'

Reverse primer: 5'-gag gta tgt gtt ctt aat caa ccc aag ttc tga gag g-3'

2.2.2.4 H-enhancer subcloning

The 2,1Kb of H-enhancer was amplified by PCR using mouse genomic DNA as template using the PWO polymerase kit with 1 cycle of 94°C for 4 minutes, 35 cycles of 94°C for 15 seconds, 65°C for 30 seconds and 72°C for 2 minutes+5 seconds/cycle, 1 cycle of 72°C for 8 minutes. PCR products were then purified with GENE CLEAN^{RII} and A-tailed with Taq polymerase for 20 minutes at 72°C by adding 200µM dATP, followed by subcloning into PGEM-T vector. The Sall/BamHI enhancer linker was generated by annealing the sense oligo strand and antisense strand for 1 minute at 95°C, 30 minutes at 65°C, 8 minutes at 37°C and 10 minutes at room temperature using ligation buffer and then ligated with PGL3-promoter restricted by Sall/BamHI. The EcoRV Restricted PGL3-promoter-linker fragment was 5'-dephosphated by Antarctic Phostase for 15 minutes at 37°C and ligated with the

2,1Kb Scal enhancer fragment extracted from the recombinant PGEM-T vector. The correct clones were identified by Sall/BamHI digestion and the orientation of the enhancer fragment was confirmed by sequencing. The sequences of the primers and the oligonucleotide for the linker are as follows:

Forward primer: 5'- ttc aca gca ttt ggg atg ttg agg-3'

Reverse primer: 5'-ctc tgc cta cca tgt gat act cca-3'

Sense strand of the linker: 5'-tcg acg ata tcg-3'

Antisense strand of the linker: 5'- gat ccg ata tcg-3'

2.2.2.5 Transcription factors subcloning

Full-length sequences of O/E-1, O/E-2, O/E-4 and Lhx-2 were obtained by amplifying plasmids containing the corresponding cDNA in 50µl of reaction solution containing 20ng of template, 200µM of dNTP, 400nM of Forward and Reverse primer, 1x reaction buffer containing 2,0mM of MgSO₄ and 0,5ul of PWO polymerase. The PCR products were purified with GENECLAN^{RII} and then treated by adding dATP to the 5'-end with Taq Polymerase for 20 minutes at 72°C in a volume of 20µl containing 100uM of dATP. The 'A-tailed' fragment was subsequently cloned to PGEM-T, and positive colonies were identified by digesting the extracted plasmid with HindIII/XhoI and further confirmed by sequencing. HindIII/XhoI fragments were then subcloned to the pcDNA3.1 (Invitrogen) expression vector. The PCR programs and the sequences of the primers are as follows:

O/E-1, O/E-2 and O/E-4: 1 cycle of 4 minutes at 94°C, 35 cycles of 15 seconds at 94°C, 30 seconds at 68°C and 1:40 minutes +5 seconds/cycle at 72°C, 1 cycle of 8 minutes at 72°C.

LHX-2: 1 cycle of 2 minutes at 94°C, 35 cycles of 30 seconds at 94°C, 30 seconds at 70°C and 40 seconds+5 seconds/cycle at 72°C, 1 cycle of 8 minutes at 72°C.

For O/E-1: 5'- aag ctt gcc gcc acc atg ttt ggg atc cag gaa agc atc-3'

Rev O/E-1: 5'- ctc gag tca cat ggg agg gac aat cat gcc-3'

For O/E-2: 5'- aag ctt gcc gcc acc atg ttt ggg att cag gag aat att ccg-3'

Rev O/E-2: 5'- ctc gag tca cat ggg cgg gac tac cag ccc-3'

For O/E-4: 5'- aag ctt gcc gcc acc atg ttc ccc gca cag gac gct ctg-3'

Rev O/E-4: 5'- ctc gag tta gga gta tgc caa gcc ctg gag-3'

For Lhx-2: 5'- aag ctt gcc gcc acc atg ctg ttc cac agt ctg tcg g-3'

Rev Lhx-2: 5'- ctc gag tta gaa aag gtt ggt aag agt gg-3'

2.2.2.6 Cell culture and transfection

HEK293 cells were cultured in DMEM medium (Invitrogen) supplemented with 10% fetal calf serum (Invitrogen), penicillin (100U/ml) and streptomycin (100µg/ml) in the incubator supplied with 5% CO₂ at 37°C. 96-welled plate (Nunclon) was employed to maintain the cells for the transfection experiments. The cells were transiently transfected with 200ng of pGL-3 based reporter plasmid along with 200ng of pcDNA3.1 plasmid expressing O/E-1, O/E-2 or O/E-4 using Lipofectamin (Invitrogen) and the transfection was performed according to the manufacture instructions. All transfections were adjusted to 500ng of total DAN with empty pcDNA3.1 vector.

2.2.2.7 Measurement of luciferase activity

The medium was removed from the plate and 20µl of lysis buffer (Invitrogen) was added to cover the cells. After rocking the plate for several times and scraping the cells from the well the suspension was transferred to a microcentrifuge tube put onto ice, followed by vortexing for 15 seconds and centrifuging for 15 minutes at room temperature by 12,000g. The supernatant was then transferred to the well of a measurement plate. Following addition of 100µl of Luciferase assay reagent the measurement was immediately initiated using a Luminescence Spectrometer (PERKIN ELMER LS50B) following the program: measurement delay for 2 seconds and measurement read for 10 seconds.

2.2.3 Bioinformatic studies

2.2.3.1 Download of genomic sequence

The mRNA or amino acid sequence of each mouse olfactory receptor gene within a cluster was obtained by referring to the corresponding gene at <http://www.ncbi.nlm.nih.gov/>. This sequence was blasted against the genome of other species and the orthology was determined according to the annotation by NCBI or Ensemble. The sequences of the conserved loci were downloaded from <http://www.ncbi.nlm.nih.gov/mapview/> or from <http://www.ensembl.org/index.html>.

2.2.3.2 Formatting of genomic sequence

Repetitive DNA elements were removed from the genomic sequence by RepeatMasker (Smit and Green., 1999) (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>) (Version: Open-3-1-7). The multi N replacing the repetitive elements was removed by Microsoft Word and subsequently the sequence was

transferred to FASTA Format using Gendoc.

2.2.3.3 Comparative analysis of genomic sequence by PipMaker

The Pipmaker program was available at (<http://pipmaker.bx.psu.edu/pipmaker/>) (Schwartz et al., 2000). To align multi sequences the program MultiPipMaker was employed (<http://pipmaker.bx.psu.edu/cgi-bin/multipipmaker>). To align two sequences for obtaining the dot matrix the program Advanced PipMaker was chosen (<http://pipmaker.bx.psu.edu/cgi-bin/pipmaker?advanced>). The default parameters were employed except in few cases "High sensitivity and low time limit" was selected.

2.2.3.4 Search for transcription factor binding sites

The sequence of the conserved segment was retrieved from the genomic sequence with program Genomics Expression and then subject to search for transcription factor binding sites using program Match-Public (<http://www.gene-regulation.com/cgi-bin/pub/programs/match/bin/match.cgi>). To limit the search the matrix database for vertebrate was selected and to minimize the false positive and negative matches "to minimize the sum of both error rates" was selected. The other parameters were set up as default settings.

2.2.4 MOR37C transgenic mouse

2.2.4.1 Plasmid construction

A 10,575 bp genomic fragment containing the *mOR262-12* coding region was generated by ligating a 785 bp segment upstream from the Spel-site to the Spel-EcoRI fragment previously used for homologous recombination (Strotmann et al., 2000). A PmeI site was generated at the 5' end of the construct for isolation. An IRES-tau-lacZ cassette (Mombaerts et al., 1996) without 'neo' was inserted into the PacI site three nucleotides downstream of the stop codon of *mOR262-12*.

2.2.4.2 Generation of transgenic mice

The purified PmeI fragment was microinjected into the pronuclei of fertilized mouse eggs, followed by transplantation into C57Bl6/J mice using standard procedures (Agrobigen, Hilgertshausen, Germany). Animals were kept at the University of Hohenheim transgenic core facility; housing conditions fulfil the animal welfare guidelines. Mice were genotyped by PCR using primers 5'-GCAGCAGTTTTTCCAGTTCCGT-3' and 5'-TCGCTCGCCACTTCAACATCAA-3' positioned within the lacZ gene to generate a 728 bp fragment. PCRs were carried out using 'MasterTaq' (Eppendorf, Hamburg, Germany) with 94°C for 4 min followed

by 35 cycles 94°C (30 sec), 65°C (30 sec), 72°C (45 sec) followed by one cycle at 72°C (8 min).

2.2.5 PCR based methods

2.2.5.1 Real-time PCR

The number of transgene copies integrated into the genome was determined by real-time PCR. For this purpose the primer pairs RT-LacZ (5'-GCAGCAGTTTTTCCAGTTCC-3'; 5'-ACATCCAGAGGCACTTCACC-3') and RT-GFP (5'-AGCAGAAGAACGGCATCAAG-3'; 5'-CTGGGTGCTCAGGTAGTGGT-3') were employed which amplified fragments from the coding region of the lacZ- and GFP gene, respectively. Using these primers under the same conditions (1 cycle: 94°C for 4 min; 30 cycles: 94°C for 30 sec, 57,8°C for 30 sec and 72°C for 30 sec; 1 cycle, 72°C for 8 min) on genomic DNA from tail biopsies of double mutant mice carrying an mOR262-12-IRES-tau-GFP allele and an mOR262-12-IRES-tau-lacZ allele (Strotmann et al., 2000) revealed the same efficiency of both PCR-reactions. For determining the transgene copy number, double mutant mice hemizygous for the transgene and heterozygous for the mOR262-12-IRES-tau-GFP allele were used; genomic DNA from tail biopsies were used as template. All reactions were run on an 'ABI Prism 7000' (Applied Biosystems, Weiterstadt, Germany) using the SYBRGreen PCR-kit according to the manufacturer's specifications (Qiagen, Hilden, Germany).

2.2.5.2 Chromosome conformation capturing (3C)

The 3C was performed according to the description by Dekker et al (2002) with slight modification. Mouse was decapitated and the olfactory epithelium (OE) was dissected out. Following homogenization for 50 times in 3ml of cold homogenization buffer(0,25M sucrose, 15mM Tris-HCl pH7,9, 60mM KCl, 15mM NaCl, 5mM EDTA, 1mM EGTA, 0,5mM spermidine, 0,15mM spermine and 1mM DTT) the cells were precipitated by centrifuging for 10 minutes at 4°C by 5,000g and then washed with 3ml of fixation buffer (10mM Tris-HCl pH7,9, 10mM MgCl₂ and 1mM DTT) 10 minutes at 4°C by 5,000g. The cell pellet was fixed with 3ml of 1% paraformaldehyde in fixation buffer for 10 minutes at room temperature with frequent mixing and the fixation was quenched by adding 0,125M of Glycine. The cells were spun down by centrifuging for 10 minutes at 4°C by 5,000g and lysed in 3ml of lysis buffer (0,25% TritonX-100, 0,5% NP-40, 10 mM EGTA, 1mM EDTA, 0,1mM PMSF and 1:500 protease inhibitor cocktail) for 10 minutes at 4°C. 20 strokes of dounce

homogenization were performed to disrupt the cell membrane and the nuclei were precipitated by centrifuging for 10 minutes at 4°C by 5,000g. The nuclei were washed with restriction buffer and then resuspended in 1ml of restriction buffer containing 0,3% SDS and incubated for 1 hour at 37°C with constant agitation, followed by adding 1,8% TritonX-100 and incubating for 1 hour at 37°C with shaking to stop SDS. The nuclei were subsequently subject to digestion by adding 500 units of restriction enzyme (NcoI) and incubating overnight at 37°C with shaking. The digestion was stopped by adding 1,6% of SDS and heating for 20 minutes at 65°C and the cleaved chromatin was diluted to 2-3ng/ul with ligation buffer. After SDS was sequestered by adding 1% of TritonX-100 and incubating for 1 hour at 37°C the ligation was initiated by adding 800 units of T4 ligase and incubating for 4 hours at 16°C. The cross-linking was reversed by adding 100µg/ml of Proteinase K and incubating overnight at 65°C and the RNAs were removed by adding 0,5µg/ml of RNase and incubating for 30 minutes at 37°C. Following Phenol-Chloroform extraction and ethanol precipitation of the DNA the PCR was set up with at least of 200ng of ligation as template and primer pairs from the mOR37C promoter and H element respectively. The programs for PCR and the sequences of primers are as follows:

the first round: 1 cycle of 94°C for 4 minutes, 30 cycles of 94°C for 30 seconds, 65°C for 30 seconds and 72°C for 1 minute, 1 cycle of 72°C for 8 minutes.

the second round: 1 cycle of 94°C for 4 minutes, 30 cycles of 94°C for 30 seconds, 62°C for 30 seconds and 72°C for 45 seconds, 1 cycle of 72°C for 8 minutes.

For 1st from 37C: 5'-ccc tgg aga gca gtg tga aac-3'

For 2nd from 37C: 5'-atc ctc agt ccc cct ttt cag-3'

Rev 1st from H: 5'-att gga tca tct gtg act cgg-3'

Rev 2nd from H: 5'-ttg acc tcc aga agc ccc ta-3'

2.2.5.3 DNA walking

DNA walking was performed to identify the integration site for the transgene of line7 as described in the manual. The first round PCR was carried out in 50µl of reaction solution containing 80ng of Line7 mouse genomic DNA, 0,2µM DW-ACP Forward primer and TSP1 reverse primer and 1x Master Mix II with hot start at 94°C for 1 cycle of 94°C for 5 minutes, 1 cycle of 42°C for 1 minute, 1 cycle of 72°C for 1 minute, 30cycles of 94°C for 40 seconds, 55°C for 45 seconds and 72°C for 1 minute, and 1 cycle of 72°C for 7 minutes. The PCR product were purified by PCR Purification kit (QIAGEN) and 20µl of reaction solution was then prepared on ice

containing 2µl of the purification of the first round of PCR product, 1µM of DW-ACPN Forward primer and TSP2 reverse primer and 1x Master Mix II. The second round PCR was performed by hot start at 94°C for 1 cycle of 94°C for 3 minutes, 35 cycle of 94°C for 40 seconds, 60°C for 40 seconds and 72°C for 1 minute, and 1 cycle of 72°C for 7 minutes. Subsequently the third round PCR was performed in 20µl of reaction solution containing 1µl of second round PCR product, 1µM of universe Forward primer and TSP3 Reverse primer and 1x Master Mix II by hot start at 94°C for 1 cycle of 94°C for 3 minutes, 35 cycles of 94°C for 40 seconds, 65°C for 40 seconds and 72°C for 1 minute, and 1 cycle of 72°C for 7 minutes. The third round PCR product was then run on 2% agarose gels and extracted with Gene Clean, followed by direct sequencing with TSP3 primer. Alternatively the fragment was dATP-tailed with Taq polymerase for 20 minutes at 72°C by adding 200µM dATP, subcloned into PGEM-T vector and sequenced with T7 or Sp6 primer. The sequences of the primers are as follows:

TSP1: 5'-cca ctg act tta ccc acc atc-3'

TSP2: 5'-tac cca cca tca aca acc tcc ag-3'

TSP3: 5'-cac cat caa caa cct cca gga tc-3'

2.2.5.4 Inverse PCR

Mouse genomic DNA was digested with NspI, NcoI, AflIII or NsiI overnight at 37°C. After inactivation by heating for 20 minutes at 65°C the enzyme was removed by phenol: chloroform extraction; the digested DNA was religated at concentration of 2-5ng/µl overnight at 16°C. The ligase was then inactivated by heating for 10 minutes at 65°C and the religation was precipitated with ethanol; 150ng of religation was employed as template for PCR. The PCR (for NcoI and NspI) was performed in 50 µl of reaction solution containing 200ng of religation, 200uM dNTP, 200nM of Forward I and Reverse primer I, 1x reaction buffer and 6U of Taq polymerase for 1 cycle of 94°C for 3 minutes, 15 cycles of 94°C for 30 seconds, 63°C-0,5°C/cycle for 45 seconds, 72°C for 3 minutes, 20 cycles of 94°C for 30 seconds, 53°C for 45 seconds, 72°C for 3 minutes and 1 cycle of 72°C for 8 minutes. The second round of PCR was then performed in 50µl of reaction solution containing 1µl of PCR product of the first round, 200µM dNTP, 200nM of Forward II and Reverse II primer, 1x reaction buffer and 6U of Taq polymerase for 1 cycle of 94°C for 3 minutes, 35 cycles 94°C for 30 seconds, 65°C for 40 seconds, 72°C for 3 minute and 1 cycle of 72°C for 8 minutes. In the case of AflIII and NsiI the first round of long range PCR was performed in 50 µl

of reaction solution containing 200ng of religation, 500 μ M dNTP, 400nM of Forward I and Reverse primer I, 1x reaction buffer and 2U of polymerase mix for 1 cycle of 93°C for 3 minutes, 35 cycles of 93°C for 15 seconds, 62°C for 30 seconds, 72°C for 8 minutes and 1 cycle of 72°C for 8 minutes. The second round of PCR was performed in 50 μ l of reaction solution containing 1 μ l of PCR product of the first round, 500 μ M dNTP, 400nM of Forward II and Reverse II primer, 1x reaction buffer and 2U of polymerase mix using the same program as the first round. PCR products could be directly sequenced with primer Forward II or Reverse II, or be subcloned into PGEM-T vector and then sequenced by primer Sp6 or T7. The sequences of primers are as follows:

The reverse primers for all PCR:

Rev1: 5'-tgc tac ctc aga gaa tct aca g-3'

Rev2: 5'-tca tct act gac cca ctg ac-3'

The forward primers for NcoI and NspI:

For1: 5'-ctc agg cat gag cgt gta aag c-3'

For2: 5'-agg aag gaa gac agg ttt ctg cc-3'

The forward primers for AflII and NsiI:

For I: 5'-cca gtc agg ctt tct ttc aca g-3'

For II: 5'-aaa cca gcc atc gcc atc t-3'

2.2.5.5 Adaptor mediated PCR

Mouse genomic DNA was digested with XhoI overnight at 37°C. After inactivation by heating for 20 minutes at 65°C the enzyme was removed by phenol: chloroform extraction; the restricted DNA was ligated with the adaptor overnight at 16°C at concentration of 30ng/ μ l. The molecule ration between genomic DNA and the adaptor was adjusted between 1:1 and 1:3. The ligase was then inactivated by heating for 10 minutes at 65°C and the religation was precipitated by ethanol and employed as template for PCR. the first round of long range PCR was performed in 50 μ l of reaction solution containing 200ng of religation, 500 μ M dNTP, 400nM of Forward I (from adaptor) and Reverse primer I (from transgene), 1x reaction buffer and 2U of polymerase mix for 1 cycle of 93°C for 3 minutes, 35 cycles of 93°C for 15 seconds, 62°C for 30 seconds, 72°C for 8 minutes+5 seconds/cycle and 1 cycle of 72°C for 8 minutes. The second round of PCR was performed in 50 μ l of reaction solution containing 1 μ l of PCR product of the first round, 500 μ M dNTP, 400nM of Forward II (from adaptor) and Reverse II primer (from transgene), 1x reaction buffer

and 2U of polymerase mix using the same program as the first round. PCR products could be directly sequenced with primer Forward II or Reverse II, or be subcloned into PGEM-T vector and then sequenced by primer Sp6 or T7. The sequences of the primer pairs and the oligonucleotides for the linker are as follows:

Sense strand for linker: 5'-ctg ctg ccc gac aac cac tac ctg agc acc cac-3'

Antisense strand for linker: 5'-tcg agt ggg tgc tca ggt agt ggt tgt cgg gca gca g

For I: 5'-ctg ctg ccc gac caa cca cta-3'

For II: 5'- ctg ccc gac aac cac tac ct-3'

Rev I: 5'-tgc tac ctc aga gaa tct aca g-3'

Rev II: 5'-tca tct act gac cca ctg ac-3'

2.2.6 Fluorescence in situ hybridization (FISH) on metaphase chromosomes

Metaphase chromosomes were prepared from mouse livers at embryonic stage 13 (E13) as described at the 'Cytogenetic Models Resource', The Jackson Laboratories (http://www.jax.org/cyto/5day_preps.html). Prior to hybridization, chromosomes were treated for 3 min at 94°C in 100% ethanol, denatured by incubation with 0,07M NaOH for 3 minutes at room temperature and then passed through a series of ethanol (70%, 90% and 100%), each for 5 min at room temperature. For the detection of the transgene a 1,2 kb fragment from the coding region of LacZ was biotin-labeled by PCR using the primers 5'-GCAGCAGTTTTTCCAGTTCCGT-3' and 5'-AGATGGCGATGGCTGGTTTC-3' and biotin-11-dUTP. For the identification of mouse chromosome 4, BAC clone #422 M18 which spans a DNA segment adjacent to the mOR262 gene cluster (Strotmann et al., 1999) was used as template for the generation of a digoxigenin-(DIG)-labeled probe by random primed labeling (Sambrook et al., 1989). For double-FISH, 5 ng of biotin-labeled and 10 ng of DIG-labeled probes were mixed with 500 ng of Cot-1 genomic DNA in 10 µl hybridization buffer containing 50% formamide, 10% dextrane sulfate and 2 x SSC. After denaturation at 80°C for 10 minutes the mixture was immediately chilled on ice, added to the slide, covered with a coverslip and sealed with rubber cement. After hybridization overnight at 60°C the coverslip was removed and the slides washed with 2 x SSC at 68°C 3 times for 10 minutes. The DIG-labeled probe was detected with an alkaline phosphatase- (AP)-coupled anti-DIG antibody (Roche, Mannheim, Germany) and 2-hydroxy-3-naphtic acid-2'-phenylanilide phosphate (HNPP, Roche) as substrate; the biotin-labeled probe was detected with horseradish peroxidase- (HRP)-coupled streptavidin and HNPPD (Roche; 1:100 in DAP-buffer,

pH 8.0) and the Tyramide Signal Amplification (TSA) kit (Perkin Elmer, MA, USA). Chromosomes were mounted in Vectashield (Vector Laboratories, Burlingame, CA) containing DAPI.

2.2.7 X-gal staining

Mice were deeply anesthetized by CO₂ and sacrificed by decapitation. For whole mount staining, the specimen were immersion-fixed for 30 min on ice with 4% paraformaldehyde (in 100 mM phosphate buffer pH 7.4, 2 mM MgSO₄, 5 mM EGTA) and stained as follows: washed with buffer A (100 mM phosphate buffer pH 7.4, 1 mM MgCl₂ and 5 mM EGTA) once for 5 min and once for 25 min at room temperature; followed by two incubations for 5 min at room temperature in buffer B (100 mM phosphate buffer pH 7.4, 1 mM MgCl₂, 0.01% sodium deoxycholate and 0.02% Nonidet P40). The blue precipitate was generated by exposure in the dark at 37°C to buffer C (buffer B with 5 mM potassium-ferricyanide, 5 mM potassium-ferrocyanide and 1 mg/ml of X-Gal). To allow for comparison of staining intensity all specimen were incubated for a period of 1 hour. For sections the tissues were fixed by immersion in 4% paraformaldehyde in PBS at 4°C for 2 hours followed by incubation in 25% sucrose overnight, embedding in 'Tissue Freezing Medium' and freezing on dry ice. 12 µm sections were generated using a Jung CM3050S cryostat (Leica). Sections were incubated for 5 min in buffer A, 5 min in buffer B followed by buffer C.

2.2.8 Immunohistochemistry

For immunohistochemistry, tissues were fixed by immersion in 4% paraformaldehyde in phosphate buffered saline (PBS) at 4°C for 2 hours, followed by incubation in 25% sucrose overnight, embedding in 'Tissue Freezing Medium' and freezing on dry ice. Cryostat sections (12 µm) were air dried for 2 hours and rinsed for 10 min at room temperature in PBS containing 0.3% Triton-X100/10% normal goat serum (NGS). Sections were incubated overnight at 4°C with a mouse anti-beta-galactosidase antibody (Promega) diluted 1:500 and a rabbit-anti-GFP antiserum (Molecular Probes) diluted 1:1000 in PBS/0.3% Tween20. After three rinses for 5 min in PBS the bound primary antibodies were visualized using goat anti-rabbit-Alexa488 and goat anti-mouse-Alexa568 conjugated secondary antibodies (Molecular Probes) diluted 1:500 in PBS/0.3% Tween20 for 2 hours at room temperature. After rinsing three times (5 min each) with PBS, sections were mounted in Vectashield mounting media and coverslipped for microscopic analysis. No staining was seen when the

primary antibodies were used to stain tissue from wildtype animals.

2.2.9 Cell counts

Transgene expressing cells were counted according to Ebrahimi and Chess (2000) on X-Gal stained whole-mount preparations of nasal cavities. The labeled cells on the medial part of the turbinates exposed towards the nasal septum were counted under a Zeiss Stemi 2000-C stereomicroscope.

2.2.10 Microscopy and photography

Sections and chromosomes were analyzed using a Zeiss Axiophot microscope and images were captured using a 'SensiCam' CCD-camera (PCO-imaging, Kelheim, Germany) or a SV MICRO sound vision camera or a Zeiss LSM 510 META confocal microscope. Whole-mount specimen was examined and photographed using a Zeiss Stemi 2000-C stereomicroscope. Pictures were adjusted for contrast and brightness using Adobe Photoshop 7.0 (Adobe Systems Inc., Saggart, Ireland).

3. Results

3.1 Functional interaction between mOR37 promoter and transcription factors in a heterologous system

3.1.1 Background

The specific function of neurons is achieved by the differential usage of stage and tissue specific genes; the expression of genes is controlled by the interaction of regulatory elements, either promoter or enhancer and transcription factors. One group of novel helix-loop-helix (bHLH) proteins, notably of O/E-1, O/E-2, O/E-3 and O/E-4, were reported to be expressed in immature and mature olfactory sensory neurons and are considered to regulate the neuronal development by controlling the expression of specific genes (Wang et al., 1993; Liberg et al., 2002; Wang et al., 1997; Wang 2002; Wang 2003). Another candidate which seems to participate in the development of OSNs is the LIM-homeoprotein Lhx-2, whose deficiency blocks the differentiation of OSN and causes the silence of some olfactory receptor genes (Hirota et al., 2004). For both O/Es and Lhx-2 mutations in their binding sites within the OR M71 gene reduced the number of OSNs expressing M71, implying a likely direct involvement of O/Es and Lhx-2 in controlling OR gene expression (Rothman et al., 2005).

The cluster of the MOR37 genes maps at position 21.5cM (42.7Mb) on band 1 of mouse chromosome 4 and comprises 5 members with high similarity of their coding sequences (Strotmann et al., 1999; Hoppe et al., 2000). Sequence analyses have identified a conservation of their proximal 5-flanking region; reflected by several highly conserved motifs within a short segment (Hoppe et al., 2003b). Further investigation using Yeast One-Hybrid Screening led to the identification of potential regulators which bind to the conserved elements. However, the functional interaction between the putative mOR37 gene promoter and the regulators is still elusive. Thus, a heterologous system was generated which would allow to examine the functional interaction between a putative mOR37 promoter and O/Es or Lhx-2 by monitoring the expression of a reporter gene. In view of the common false positive results the promoters of other OR genes were also included as the control.

3.1.2 The strategy for cotransfection

The PGL-3 basic plasmid lacking an internal promoter was used to insert the respective putative OR gene promoter upstream the cDNA encoding the luciferase

reporter gene. To express O/Es and Lhx-2 the amplified coding sequences were subcloned into the expression vector pcDNA3.1 (Figure 1). If regulators and putative promoters interact, a cotransfection of PGL-3 and pcDNA3.1 based plasmids into HEK cells would cause an increased expression of the reporter gene and high luciferase activity compared to a combination of PGL-3 based plasmid and empty pcDNA3.1.

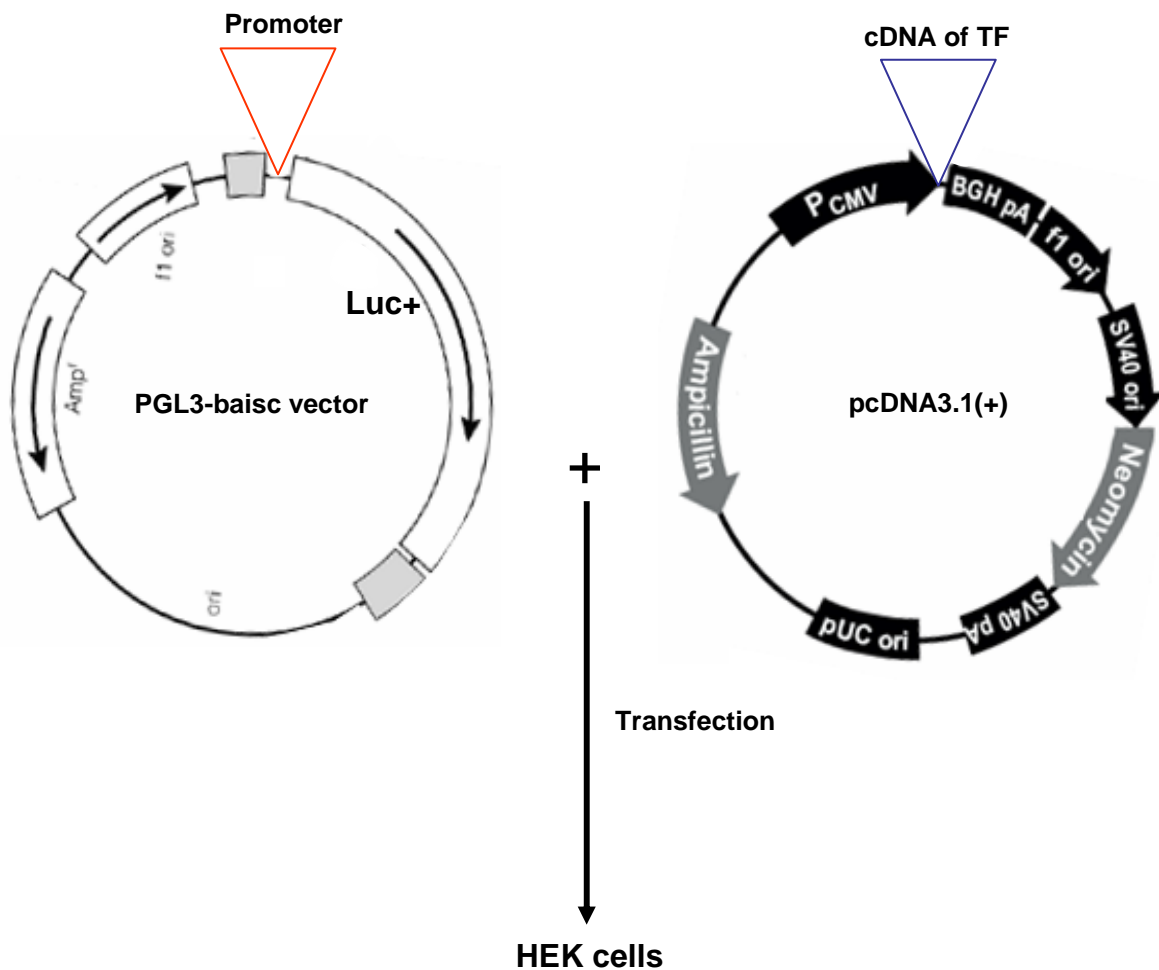


Figure 1 Outline for the generation of recombinant plasmids and the following cotransfection. The promoter segment indicated by red triangle was inserted into pGL-3 basic vector to drive the expression of luciferase gene (Luc+) and the cDNA of TF (Transcription factor), shown by blue triangle, was subcloned into pcDNA3.1 vector for expression driven by the CMV promoter. HEK293 cells were transfected by the mixture of the plasmids with Lipofectamin.

3.1.3 Monitoring the interaction of Lhx-2 and putative mOR37 promoters by luciferase expression

Cotransfection of the mOR37A or mOR37B promoter with Lhx-2 resulted in an increase of the luciferase activity by 4.1 ± 0.6 fold and 4.68 ± 0.46 fold, respectively. Cotransfection of the mOR37C promoter and Lhx-2 did not lead to significant changes. This was also the case for other promoters such as mOR120-1. Cotransfection of the class I receptors mOR18-2 and mOR11-1 with Lhx-2 led to an enhanced luciferase activity by 5.85 ± 0.49 fold and 2.04 ± 0.31 fold, respectively (Figure 2). Comparing the number of homeodomain sites of different promoters (Table 1) revealed that there is no direct relationship between the degree of activation and the number of binding sites.

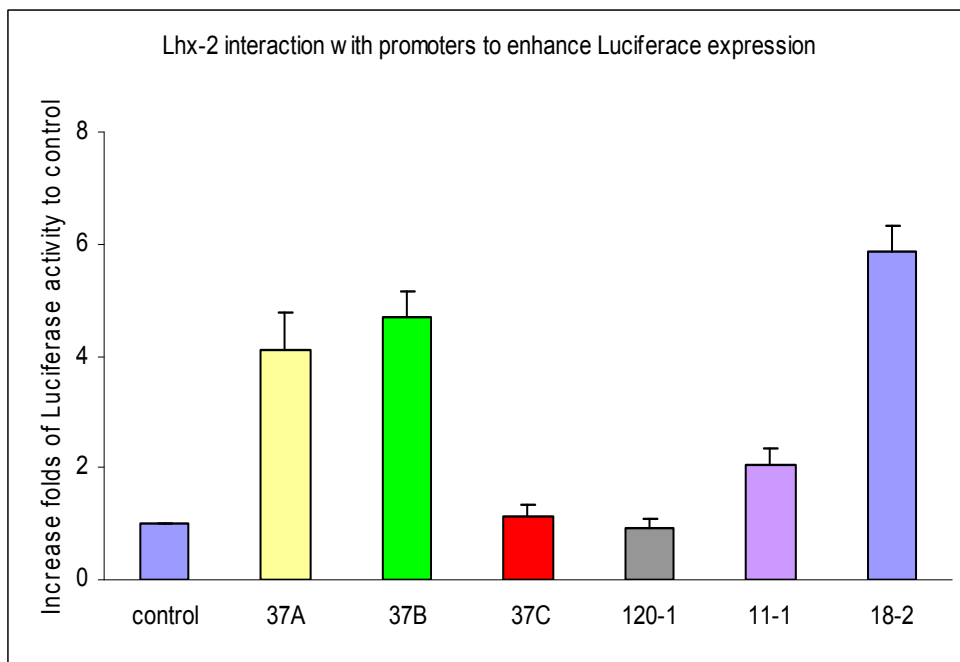


Figure 2 Interaction between Lhx-2 and the putative mOR promoter affects the expression of luciferase gene in HEK-293 cells. Cotransfection of pcDNA empty vector and promoter fused PGL-3 plasmid was employed as control. Enhanced luciferase expression was determined as the ratio of luciferase activity from cotransfection of Lhx-2 (cDNA)-pcDNA and promoter-PGL-3 plasmid to that from the control. The putative promoter of the various OR-genes are indicated at the X axis. The data are the results of six independent experiments.

OR promoter	37A	37B	37C	120-1	11-1	18-2
Number of site	2	6	1	6	2	5

Table 1. Number of homeodomain sites in different promoters.

3.1.4 Site-directed mutagenesis of bases flanking the Homeodomain-like site in the promoter mOR120-1

In order to elucidate the basis for the different responsiveness of the related promoters, attempts were made to modify the sequence composition of regions flanking the binding sites. Therefore, the bases flanking the Homeodomain sites in the various promoters were carefully examined and compared; it turned out that in all promoters which displayed no functional interaction, the Homeodomain sites was flanked by the same base. In contrast, promoters which showed a functional interaction with Lhx-2 possessed at least one Homeodomain site flanked by different bases (Figure 3). Thus, it was assumed that a conversion of the flanking base may enable the promoter to functionally interact with Lhx-2. Therefore, site directed mutagenesis experiments were performed changing in promoter mOR120-1 the upstream base of the third Homeodomain site from T to C. Subsequent luciferase assays showed no evidence for any enhanced activity; indicating that there was still no functional interaction (Figure 4). To unravel the mechanisms accounting for the functional interaction between Lhx-2 and the Homeodomain sites further analyses are needed.

```

mOR37A:  ...ataatt...
mOR37B:  ...ttaatt..., ...ctaatt...
mOR37C:  ...ttaatt...
mOR120-1: ...ctaattc..., ...ttaatt..., ...Ttaatt...
mOR11-1: ...ctaatt..., ...ttaata...
mOR18-2: ...ctaattg..., ...ataatt..., ...ataatt..., ...ataatc, ...ttaatc...

```

(c)
↑

Figure 3 Comparison of Homeodomain sites from different promoter sequences. Blue indicates the different flanking bases and red shows the identical bases flanking the Homeodomain sites. The bold red T shows the base site changed by mutation.

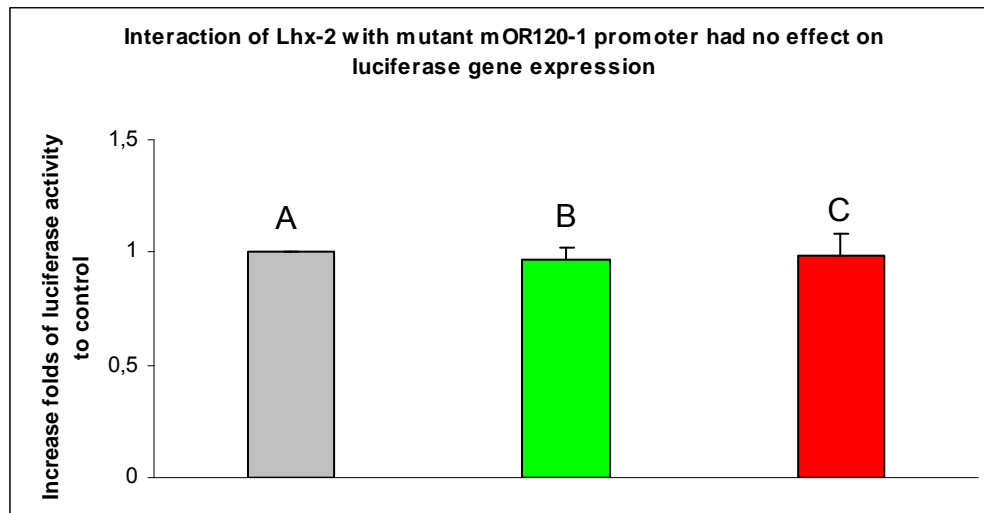


Figure 4 Luciferase activity induced by a mutated OR120-1 promoter.

A: control: cotransfection of mutant mOR120-1 and empty pcDNA3.1 vector.

B: Ratio of cotransfection of mOR120-1 and empty pcDNA3.1 vector to control.

C: Ratio of cotransfection of mutant mOR120-1 and Lhx2- pcDNA3.1 vector to control.

The data are the results of three independent experiments.

3.1.5 Functional interaction of O/E-factors with different promoters

In vitro and *in vivo* experiments have indicated that factors of the O/E family may play an important role in governing the expression of olfactory specific genes, such as OMP and ACIII (Wang et al., 1997; Robert et al., 1997; Wang et al., 2002; Wang et al., 2003). However, only in one case, evidence for a regulation of receptor expression by O/E-s was reported (Rothman et al., 2004). The question arises whether expression of OR genes may be regulated by direct interactions between the O/E-s and the promoters. To approach this aspect, two members of the mOR37 gene family, mOR37A and mOR37B, were investigated; it was found that the presence of O/E-s lead to an enhanced expression of the promoter-controlled expression of a reporter gene. For the mOR37A promoter the activity increase was determined for the O/E-1 as 2.93 ± 0.60 ; for the O/E-2 as 4.12 ± 0.87 and for O/E-4, 3.8 ± 0.5 (Figure 5). For the mOR37B promoter the activity increase was determined for the O/E-1 as 5.65 ± 0.78 ; for the O/E-2 as 5.26 ± 0.29 and for the O/E-4 as 5.17 ± 0.87 (Figure 6). These results suggest that all O/E-s may have similar effects on the expression of mOR37 OR genes. To confirm this notion the mOR37C gene was investigated. It

was found that the mOR37C promoter exhibited a different response spectrum compared to mOR37A and mOR37B, both in the extent of activity increase by the O/Es and also concerning the relative effectiveness of distinct O/Es. The combination of mOR37C and O/E-1, mOR37C and O/E-2 and mOR37C and O/E-4 enhanced the luciferase activity by $13,83 \pm 1,66$ fold, $21,70 \pm 4,39$ fold and $9,77 \pm 1,30$ fold respectively, which was about 3-5 fold stronger than mOR37A or mOR37B (Figure 7). Unlike the effect on mOR37A or mOR37B, the individual O/E subtypes showed distinct functional interaction for mOR37C.

To pursue the notion the O/Es may be involved in controlling the cluster-like expression of OR37 genes, the promoter of the mOR120-1 gene was investigated which also shows the same cluster-like expression pattern (Pyrski et al, 2001; Hoppe et al, 2006.). It was found that all O/Es had a quite similar effect: O/E-1, $8,93 \pm 1,13$; O/E-2, $9,36 \pm 0,95$; O/E-4, $7,19 \pm 1,33$ but the extent of increase in luciferase activity was higher than for the mOR37A and mOR37B promoter but less than that for mOR37C promoter (Figure 8).

Subsequently, two class I receptor genes were investigated. It was found that the promoters of mOR11-1 and of mOR18-2 could also drive the luciferase expression by interacting with all the O/Es but only to a small extent (Figure 9A and B). The summary of the effect of all O/E proteins on the putative promoters under investigation is outlined in Figure 10; it clearly shows a variability of O/E proteins enhancing luciferase activity for distinct promoters. Comparing all the promoters for O/E-1 like sites it was found that in most promoters several sites exist; only the mOR37C promoter appeared to have only a typical Olf-1 site. Whether this typical site accounts for the strong activation by O/Es is unclear.

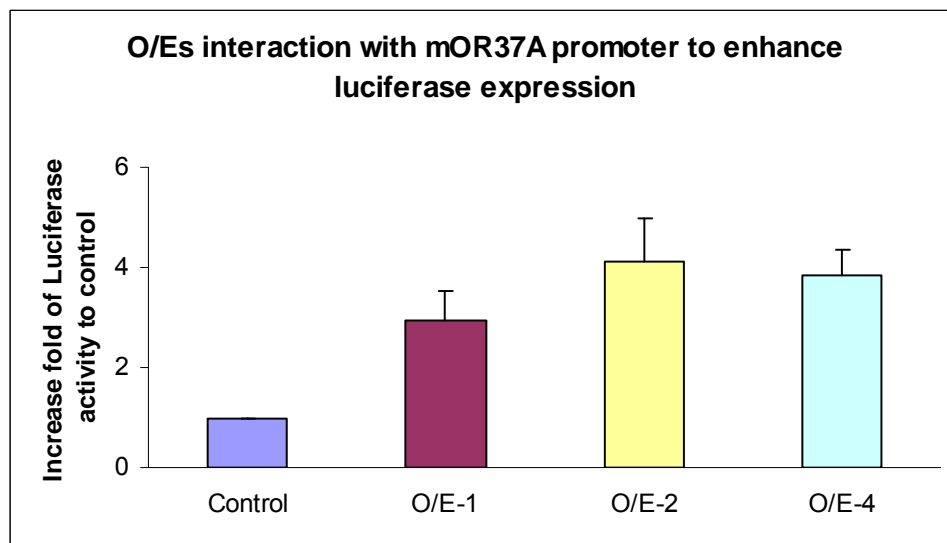


Figure 5 Functional interaction between O/E-s and mOR37A promoter. The results are from six independent experiments.

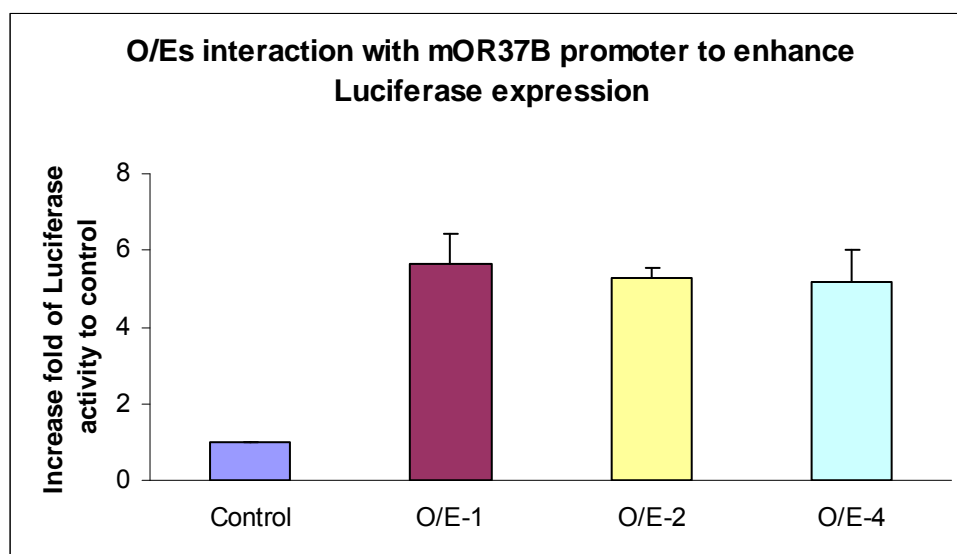


Figure 6 Functional interaction between O/E-s and mOR37B promoter. The results are from six independent experiments.

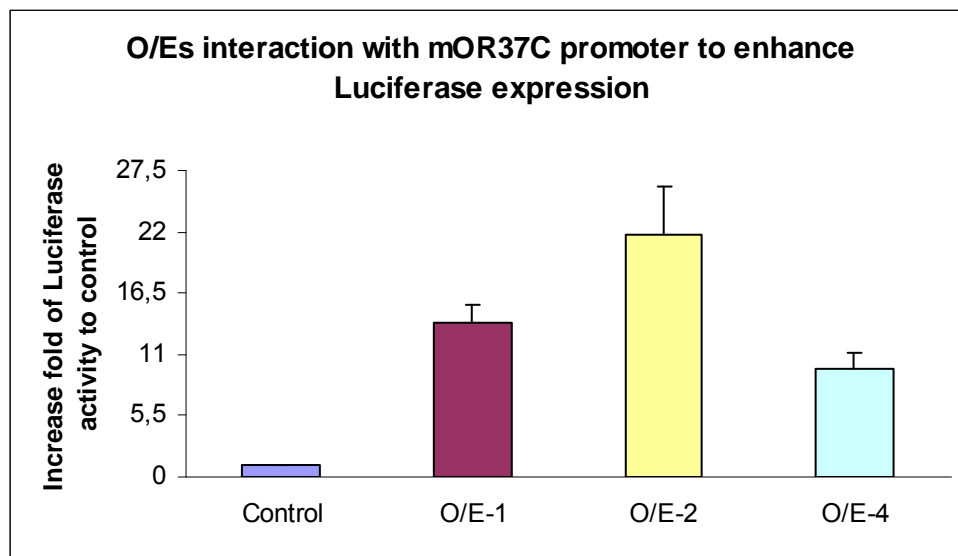


Figure 7 Functional interaction between O/Es and mOR37C promoter. The results are from six independent experiments.

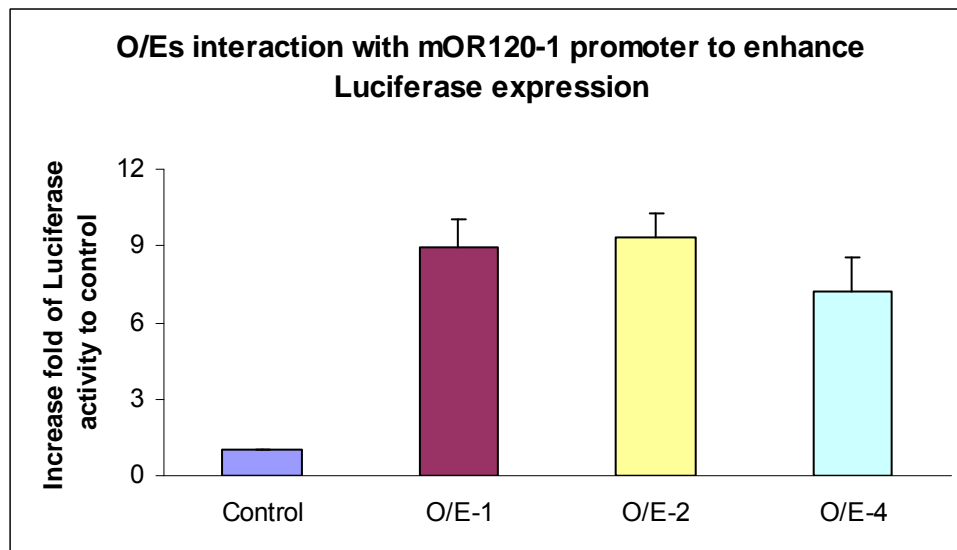


Figure 8 Functional interaction between O/Es and mOR120-1 promoter. The results are from six independent experiments.

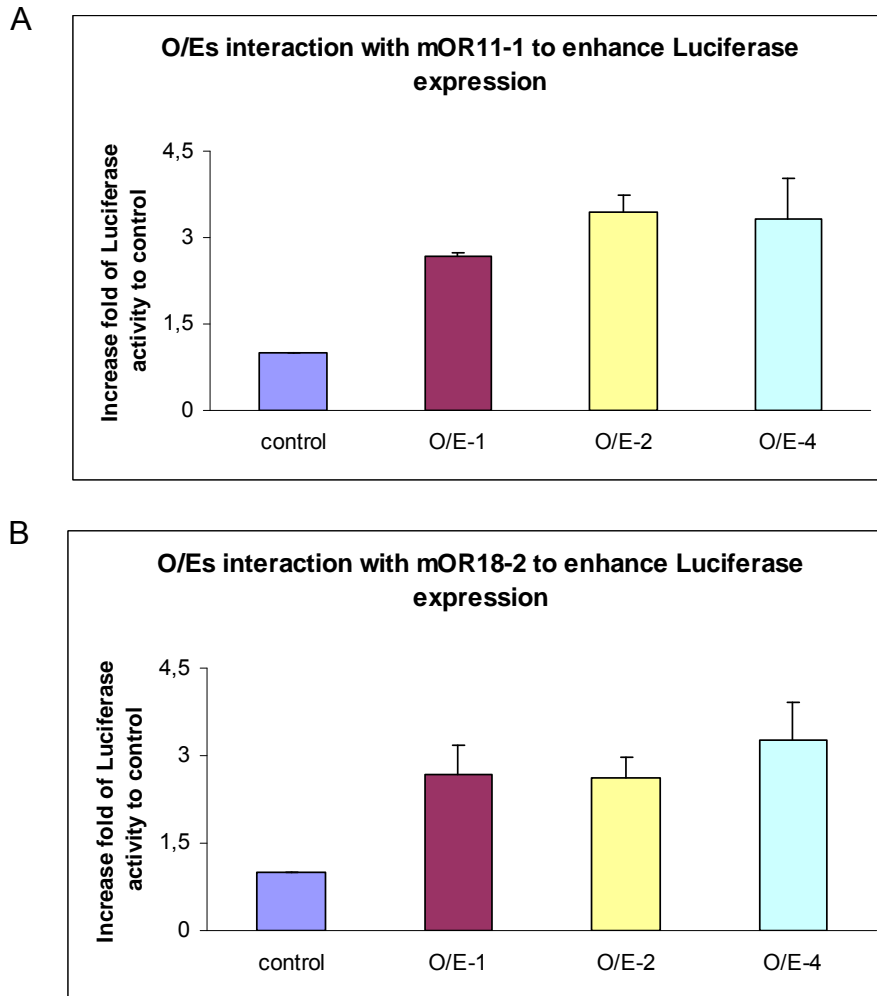


Figure 9 Functional interaction between O/Es and promoters of Class I OR genes. The results are from six independent experiments.

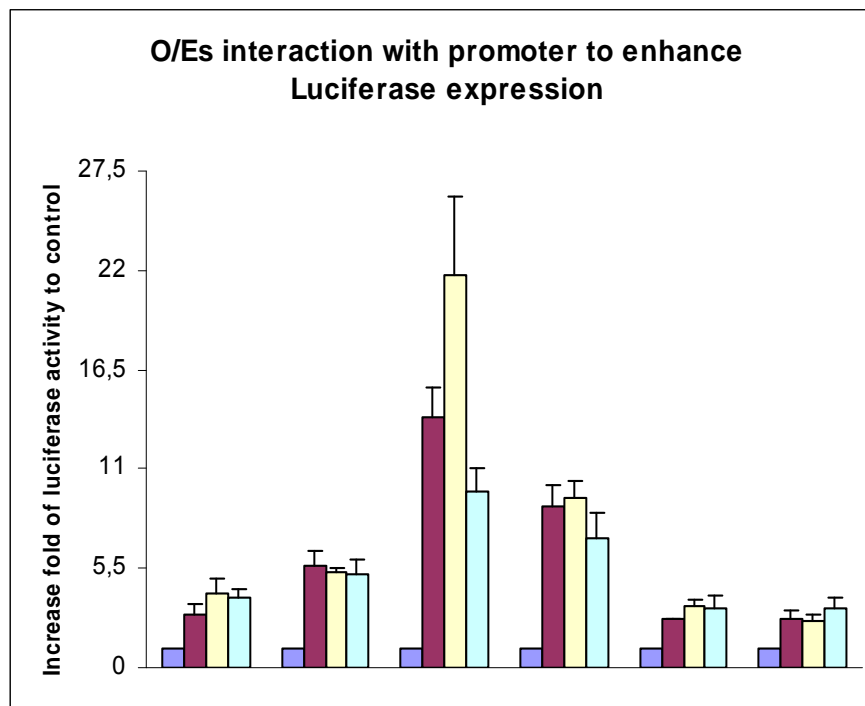


Figure 10 Summary of functional interaction between O/Es and promoters. ■ : O/E-1; □ : O/E-2; □ : O/E-4. The results are from six independent experiments.

3.1.6 Site directed mutagenesis of the olf-1 site in the mOR37C promoter

The typical olf-1 site in the mOR37C promoter is located 180 bases upstream of the transcription start site (TSS) and is composed of 12 bases: ATCCCAGGAGGA (Figure 11). The three core bases CCC were changed into GAA by site directed mutation; subsequently the reactivating effect of O/Es on the promoter was monitored by determining the expression of luciferase. It was found that in the presence of the WT mOR37C promoter + empty pcDNA3.1 or in the presence of mutant mOR37C promoter+ empty pcDNA3.1 the luciferase activity showed no difference. Therefore, the combination of mutant mOR37C promoter + empty pcDNA3.1 was used as control. The assays revealed that with the modified olf-1 site only a slight increase of luciferase activity was observed in the presence of O/Es. For O/E-1 the increase was $1,3 \pm 0,0$; for O/E-2 it was 2.0 ± 0.6 and for O/E-4 it was 1.8 ± 0.6 fold (Figure 12). Compared to the intact mOR37C promoter the activity was reduced by about 10 fold for each of the O/Es. Thus, a distinct structure of the binding site within the mOR37 promoter plays the crucial role in mediating the functional interaction between promoter and transcription factor.

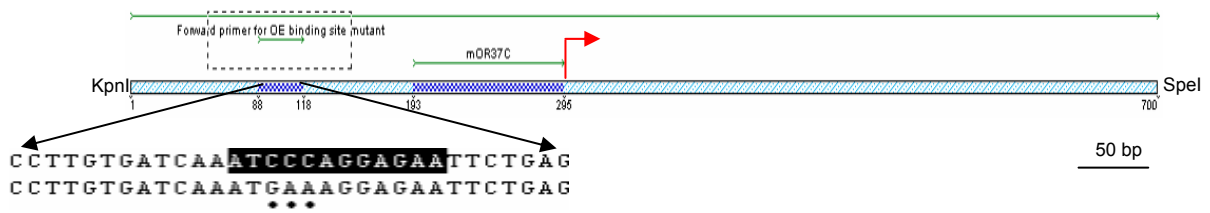


Figure 11 The strategy for site directed mutation of the olf-1 site within the promoter of mOR37C. 700 bases of a Spel/KpnI fragment is the mOR37C promoter region subcloned into PGL-3 basic vector, shown by the upper green arrow in the alignment. The red arrowhead indicates the transcription start site (TSS) and the green arrow with mOR37C above it shows the conserved element among the analyzed genes. The alignment between the mutation primer (the smallest green arrow) and the promoter shows the position and sequence of the primer. The typical olf-1 site is enclosed by the black square block and the 3 mutant bases are indicated by the black dots beneath the alignment.

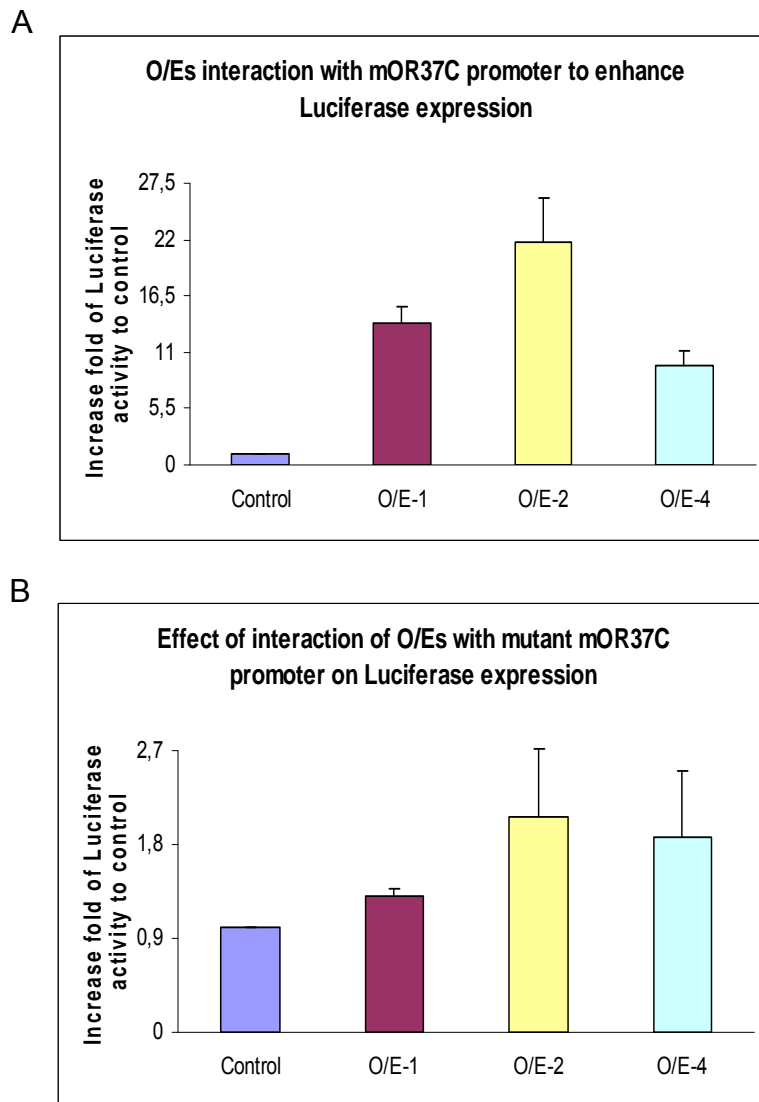


Figure 12 O/E-s induced transcriptional activity mediated by the unmodified (A) and the mutated mOR37C promoter (B). The results are from six independent experiments.

3.1.7 Simultaneous activation of the mOR37B promoter by O/E-2 and Lhx-2

Since both Lhx-2 and O/E-2 displayed positive effects on mOR37 promoters the question arises whether the interaction of distinct transcription factors may interfere with each other with negative or positive consequences. Therefore, experiments were performed monitoring the simultaneous effect of O/E-2 and Lhx-2 on the mOR37B promoter. A vector harbouring the mOR37B promoter was used to transfect HEK cells together with an Lhx-2 containing plasmid and an O/E-2 containing plasmid. Subsequently, the luciferase activity was determined. It was found that Lhx-2 or O/E-2 alone induced an increase of $4,21 \pm 0,25$ and $5,34 \pm 0,96$ fold, respectively (Figure 13). In the presence of both transcription factors an enhancement of 8.45 ± 1.80 fold was measured (Figure 13). Comparing this result and the sum of the two independent effects of Lhx-2 and O/E-2 ($9,67 \pm 1,07$) revealed that Lhx-2 and O/E-2 exhibited an additive effect.

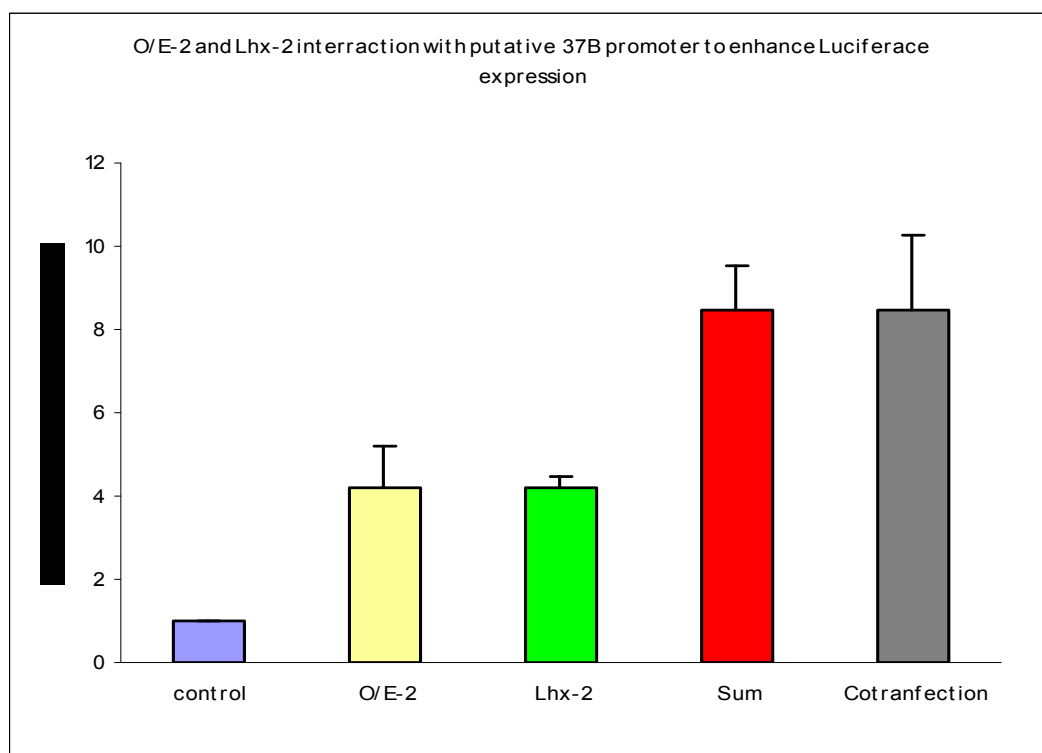
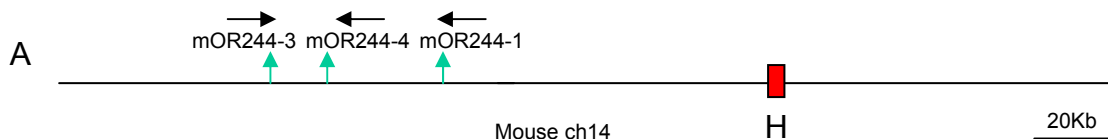


Figure 13 Enhancement of luciferase expression of a mOR37B promoter by O/E-2 and Lhx-2. The results are from six independent experiments.

3.1.8 Effect of the H element on O/E-2 interaction with mOR37B and mOR120-1 promoters

The H element is a 2kb segment located 75kb upstream of the murine mOR28 (mOR244-1) gene; it was identified by genomic sequence comparison between mouse and human (Serizawa et al., 2003). It was demonstrated that the H element activates the mOR28 gene cluster and allows the expression of only one OR gene from the cluster (Serizawa et al., 2003). Further investigation suggested that the H element might interact with multiple mOR promoters and that such a function may contribute to the “one neuron-one OR” expression through mutually exclusive interaction with only one distinct OR promoter (Lomvardas et al., 2006). In view of this concept attempts were made to examine if the H element may affect the activity of the promoter/transcription factor complex. Therefore, the H element was subcloned upstream of the mOR37B promoter and of the mOR120-1 promoter in a pGL-3 vector, which was then used to transfect HEK cells. Subsequent analysis of the cells revealed that the O/E-2 induced luciferase activity was reduced by about 2,6 fold for both the mOR37B and the mOR120-1 promoter (Figure 14). These findings indicate that under the conditions of these experiments the H element region has an inhibitory effect on the transcription activity of OR genes.



B

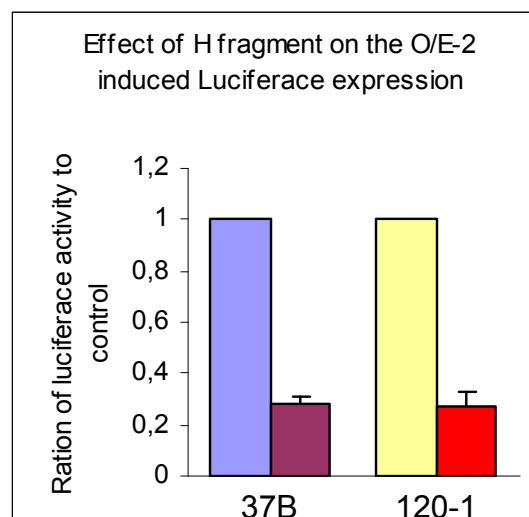


Figure 14 Reduction of O/E-2 functional interaction with mOR37B and mOR120-1 promoter by upstream insertion of the H element. A the position of H element on mouse chromosome 14. B O/E-2 triggered luciferase activity for mOR37B and mOR120-1 promoter. ■: control 37B ■: 37B+H; ■: control 120-1 ■: 120-1+H. The results are from 3 independent experiments.

3.2 *In vivo* demonstration of the role of the promoter in regulating the topographic expression of mOR37 gene

3.2.1 Expression of mOR37C transgene

The odorant receptor mOR37C (*mOR262-12*; *olfr157* according to the Mouse Genome Informatics nomenclature) is a member of family mOR262; the genes encoding these ORs form two distinct gene clusters on chromosome 4 (Strotmann et al., 1999; Hoppe et al., 2000; Hoppe et al., 2003a); mOR37C is the central gene in cluster-I, the surrounding genomic DNA thus is completely part of that particular cluster. To build the transgene a 10,5 kb genomic fragment was used (Figure 15A), which extends 5,468 basepairs (bp) downstream and 4,134 bp upstream of the coding region of *mOR262-12*. The construct includes only 358 bp upstream of the transcription start site (TSS); these contain a ~120 bp DNA-element (asterisk) which is highly conserved among all OR genes expressed in clustered OSN populations (Figure 15B) (Hoppe et al., 2006a). An IRES-*taulacZ* cassette (Mombaerts et al., 1996) was inserted three nucleotides downstream of the stop codon allowing to visualize the cells which express the transgene. After microinjection of the DNA construct into pronuclei, seven founder mice were obtained that were employed to generate independent lines (#1 - #7).

In order to determine whether the transgene was expressed in the olfactory epithelium (OE), we first performed X-Gal staining of whole-mount nasal preparations. Labeled cells were indeed detectable in the OE in all seven lines; figures 16A–C show representative examples. In lines #1 - #6 the labeled cells were visible only in the central region of endoturbinates II facing the nasal septum. This position resembled the typical clustered pattern of OSNs expressing mOR37C from the endogenous genomic locus (Figure 16D) (Strotmann et al., 2000). In line #3 only very weak signals were obtained compared to the endogenous gene. Only in one line (#7) a different pattern emerged (Figure 16E): not only was the number of cells expressing the transgene higher than in the other six lines, in addition cells were not all concentrated in the center of the turbinates but were also visible on adjacent turbinates. However, they were not randomly scattered throughout the epithelium, but formed a band along the anterior-posterior axis of the turbinates, excluding the most dorsal and ventral part of the epithelium. This pattern closely resembled the distribution of cells expressing OR genes like e.g. *P2* (Mombaerts et al., 1996)

(Figure 16F). The staining intensity of cells in line #7 appeared higher than in the other transgenic lines and also the cells expressing the endogenous gene.

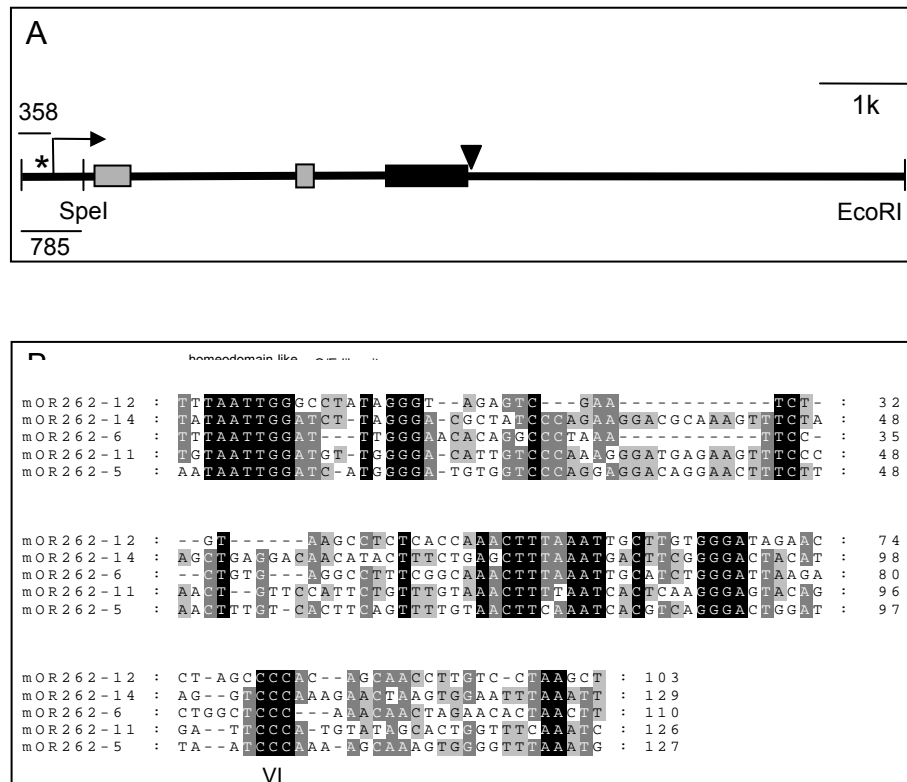


Figure 15 Construct of the transgene. (A) presents the genomic DNA fragment (10,575 bp) used for mOR37C (*mOR262-12*) transgene construction; the putative transcription start site (TSS; hooked arrow) is located 3,770 bp upstream of the mOR37C (*mOR262-12*) coding sequence (black box). Two non-coding 5' exons (grey boxes) are separated by intron sequences. The construct includes 358 bp upstream of the TSS, containing the ~120 bp DNA-element (asterisk;) which is highly conserved among all OR genes expressed in clustered OSN populations (B). The triangle indicates the insertion site three nucleotides downstream of the stop codon for an IRES-tau-lacZ cassette. Comparison of promoter sequences for representative mOR262 family genes (modified according to Hoppe et al., 2006). A conserved homeodomain site and multiple O/E-like sites are found in the region immediately upstream of the TSS.

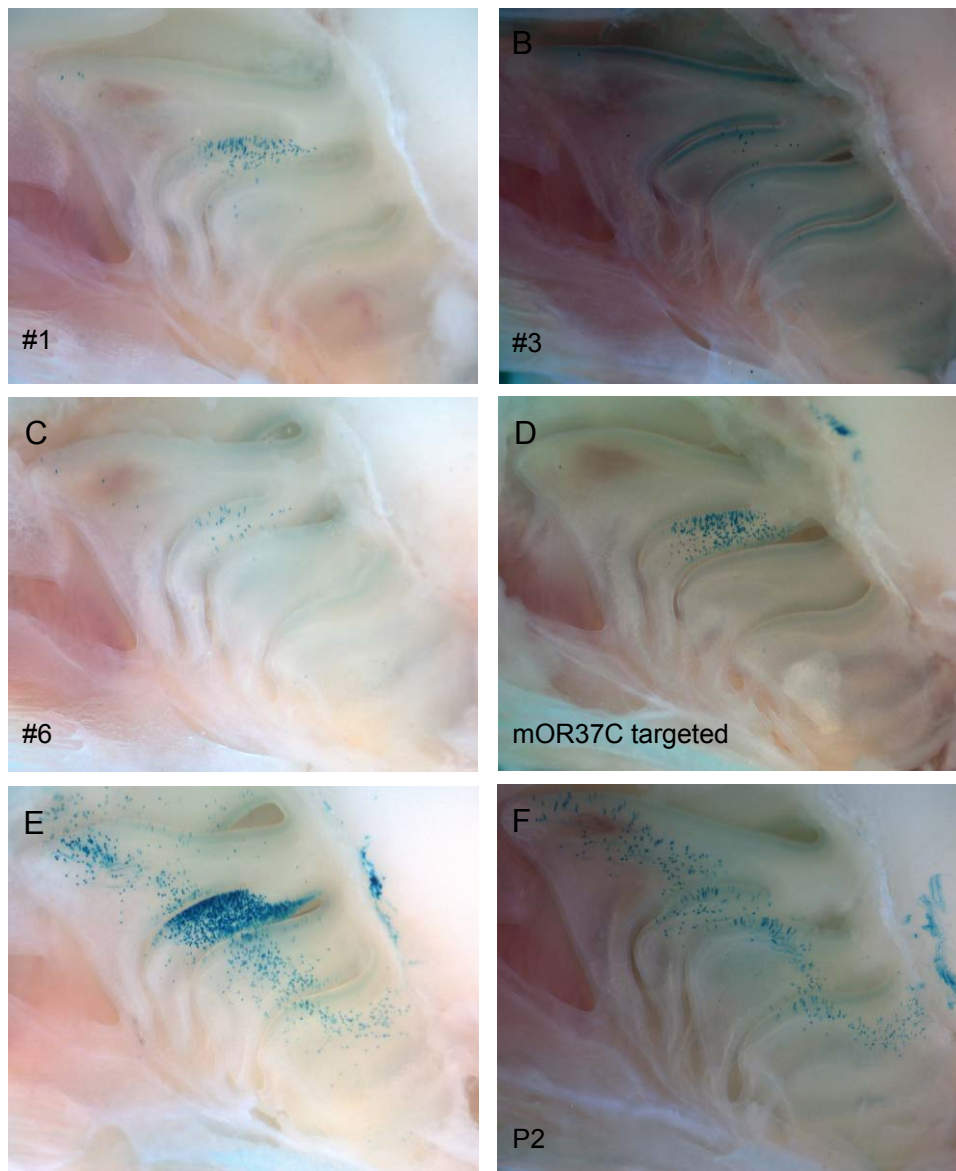


Figure 16 Expression patterns in mOR37C (*mOR262-12*) -lacZ transgenic (lines # 1,3,6,7) and targeted mice. Whole-mount preparations of sagittally transected mouse heads, showing the medial aspect of the turbinates (anterior to the left, dorsal to top). The tau-lacZ marker was revealed as a blue precipitate by incubation with Xgal. Cells expressing the transgene in lines #1- #6 (A-C) are clustered comparable to those in the targeted mouse line (D). The cell number varies among the different lines. Cells expressing the transgene in line #7 are also found outside the typical central cluster in a zonal pattern (E). (F) Whole-mount preparation of a P2-lacZ mouse head showing the typical zonal patterning of OR expressing cells.

Since whole-mount preparations allowed a view only onto the medial side of the turbinates, we next analyzed the distribution of transgene expressing cells on cross sections. As shown for example by line #1 (Figure 17A, B), transgene expressing cells in lines #1 - #6 were restricted to the tips of endoturbinates II and III and

ectoturbinate 3; this pattern is again reminiscent to OSNs expressing the endogenous gene (Strotmann et al., 2000). In line #7 the cells were more broadly dispersed over several turbinates and the septum, forming a semi-circular ring in the two nasal cavities (arrows in Figure 17C; Figure 17D). At higher magnification it became obvious that transgene expressing cells in all lines were located exclusively in the cellular layer of the OE and displayed the typical morphology of mature OSNs (Figure 18A, B). We have previously shown that in mature mice OSNs expressing a particular subtype from the mOR37 subfamily are arranged in a particular layer of the OE (Strotmann et al., 2000). On cross sections it became obvious that also the cell bodies of the transgene expressing OSNs were preferentially positioned within a rather apical position of the OE (Figure 18A, B). Interestingly, this was also found for those cells in line #7 expressing *mOR262-12* in positions outside the typical central cluster (Figure 17C, D).

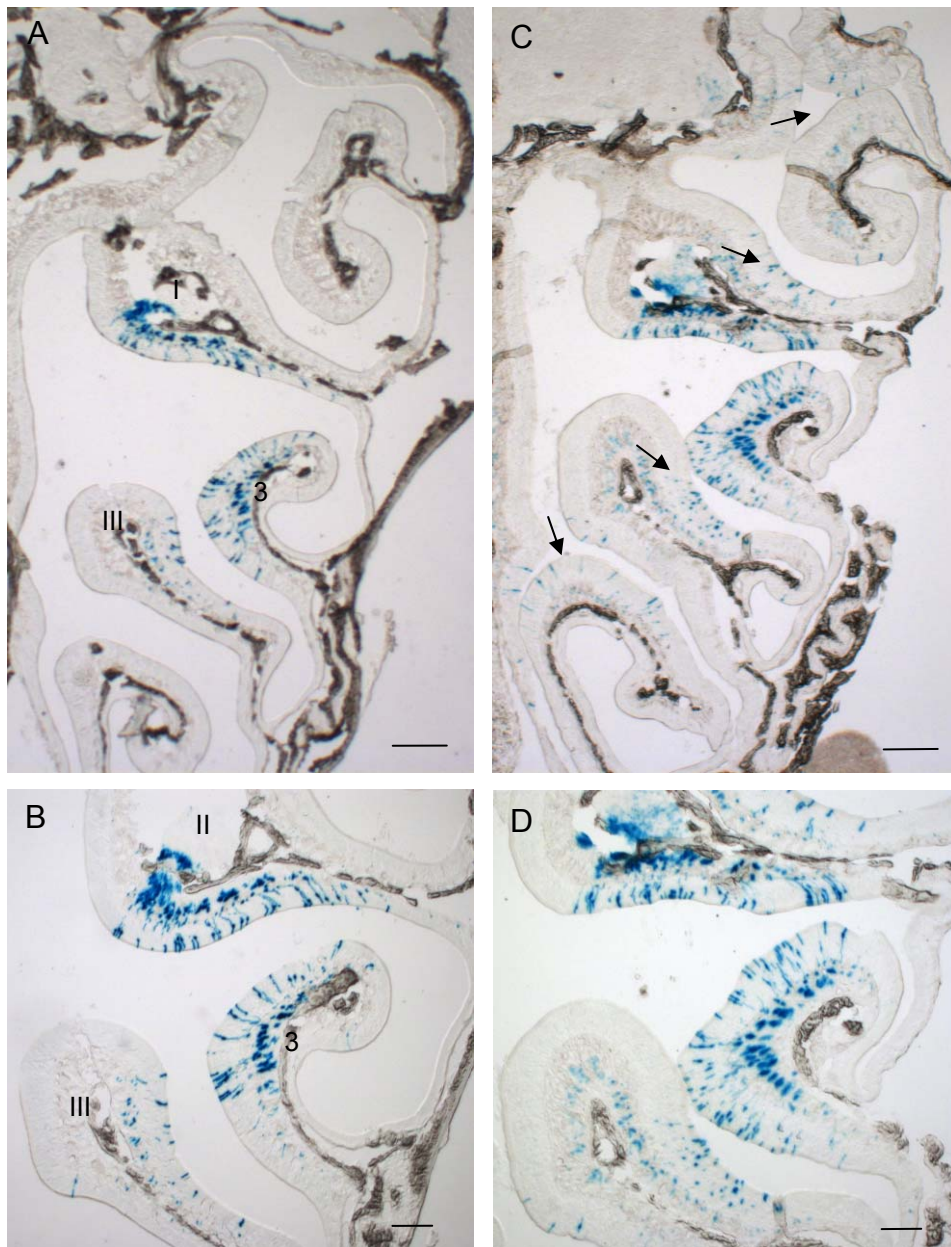


Figure 17 Microview of the transgene expression pattern on the turbinate. Pattern of transgene expressing cells in line #1 (A,B) and line #7 (C,D) shown on cross sections through the nasal cavity by incubation with Xgal. In line #1 the cells are arranged in the typical clustered pattern in three areas facing each other on endoturbinates II, III and ectoturbinate 3 (A,B). In line #7 the transgene expressing cells are found in the cluster and additionally on the septum and turbinates ranging from dorsal to ventral (arrows in C). Scale bars: 100 μ m

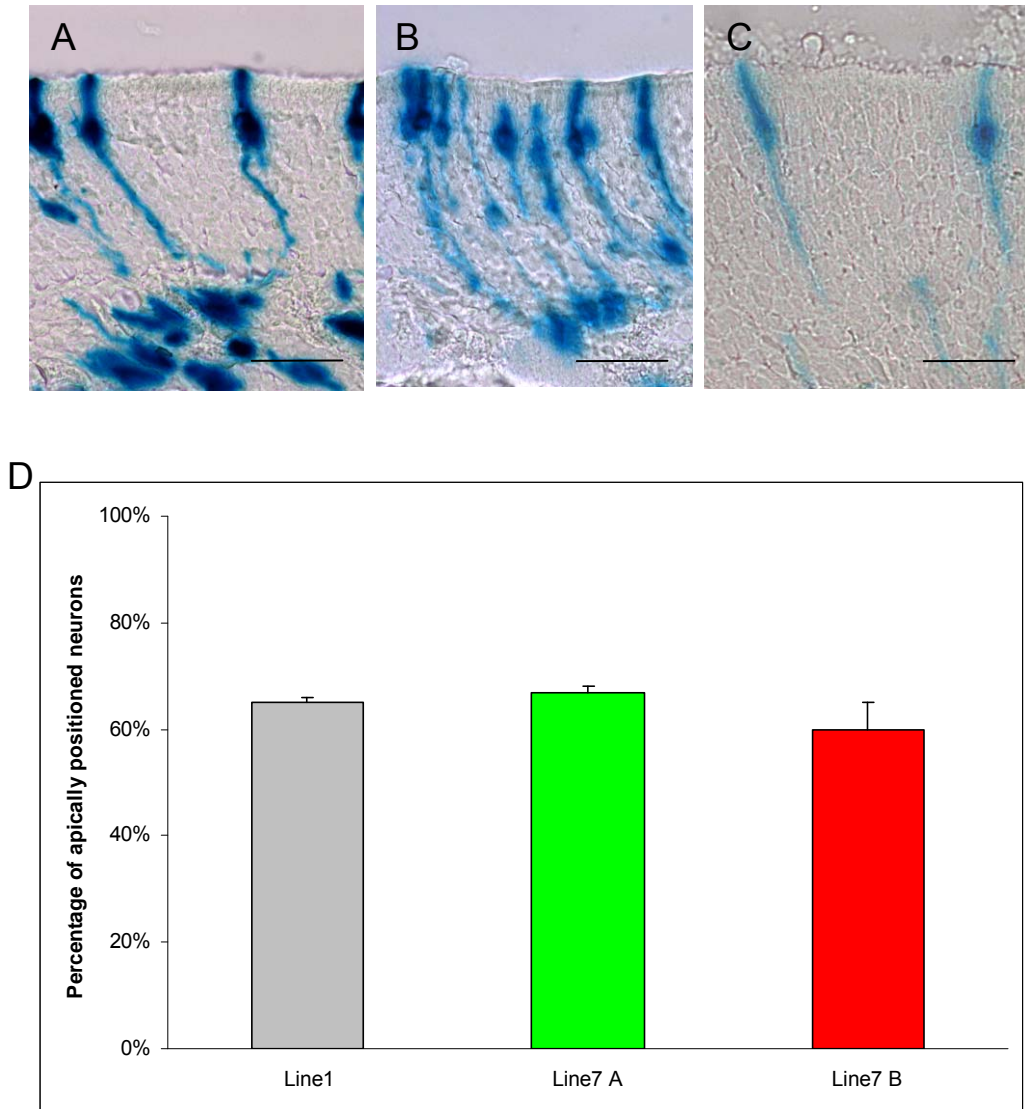


Figure 18 Apical position of the transgene expressing cells. Somata of transgene expressing cells are predominantly located in an upper layer of the epithelium in line #1 (A) and #7 (B, C). (D) The percentage of the apically positioned neurons within the normal cluster or ectopic region in transgenic mice. The apically positioned cells are defined as the ones whose a/b is within 0.8 To 1.0. a : the vertical distance of a cell body from the basal membrane, b : the distance from the basal Membrane to the nasal lumen (Strotmann et al., 2000). Line7 A: cluster Line7 B: ectopic region. Scale bars: 100 μ m.

3.2.2 Tissue specificity of transgene expression

In order to determine the tissue specificity of transgene expression, newborn mice from lines #1, #3, #6 and #7 were examined analyzing serial sections through the entire body (Figure 19A). X-Gal staining and careful inspection of all sections for labeled cells throughout all organs revealed that transgene expression occurred exclusively in the OE (Figure 19B,C and D); even few cells in the nasal epithelium of line #3 could unequivocally be identified (Figure 19C), suggesting that no other cell type in the mouse showed a “leaky” expression of the transgene.

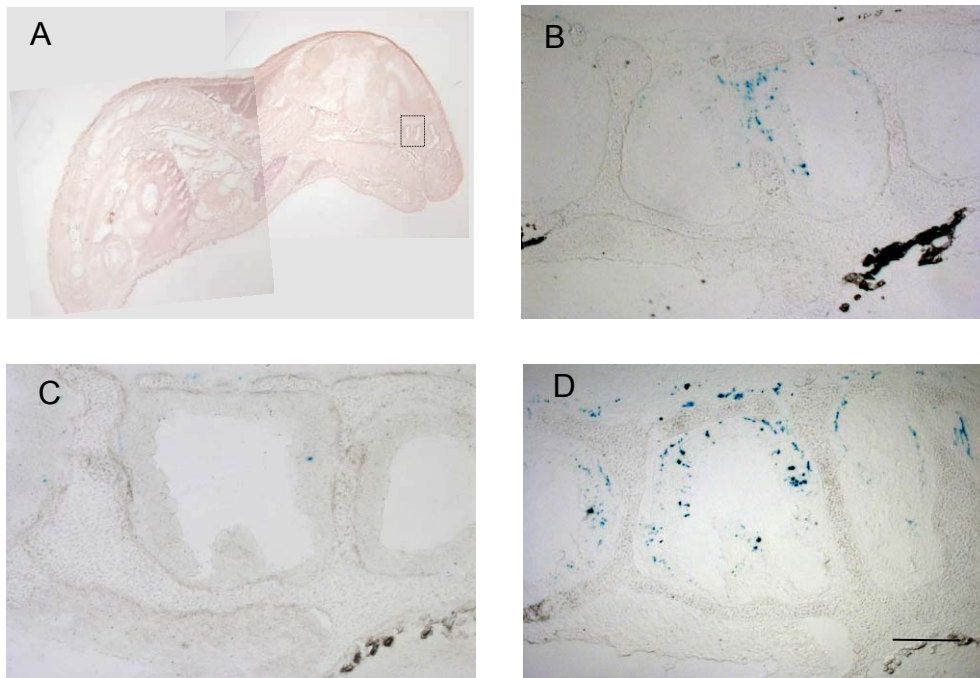


Figure 19 Specificity of transgene expression on the olfactory epithelium. Parasagittal section through the nasal cavity of lines #1 (B), #3 (C) and #7 (D) mice stained by incubation with Xgal. Transgene expressing cells in the olfactory epithelium and their axonal bundles are visible. Scale bar: 500 μ m.

3.2.3 Slight delay of the onset of transgene expression

Previous study from Strotmann et al demonstrated the mOR37C gene expression was initiated at around E11.25 (Strotmann et al., 1995). To examine whether this was the same for the transgene coronal sections of E11 to E12 embryos from the targeting and the transgenic lines were treated with X-gal staining. As shown in Figure 20A the staining was visible at the stage between E11,25 and E11,5 for the targeting mouse but no any staining could be observed at this stage for all the transgenic lines. The initial staining appeared for the transgenic lines only when the development reached the stage later than E11,5 but earlier than E12 (Figure 20B, C and D).

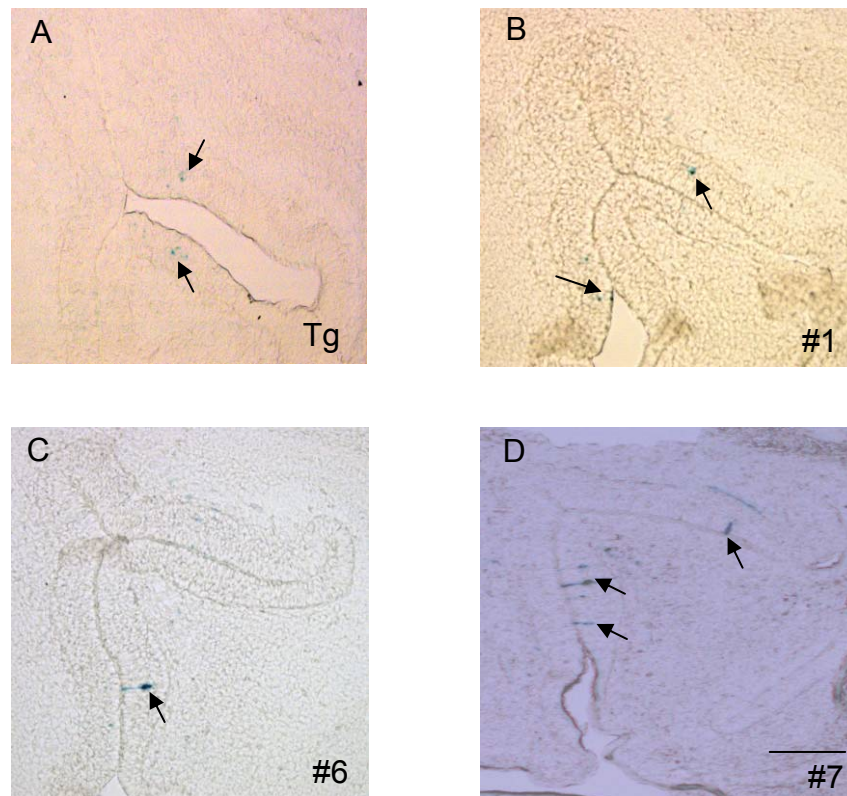


Figure 20 Onset of transgene expression at the developmental stage between E11,5 and E 12. The targeting mouse initially express the reporter gene at the stage slightly earlier than E14 (A), however, this process is started later than E11,5 but earlier than E12 in the transgenic lines (B,C,D). The orientation of the coronal sections is shown in (A): D, dorsal ; L, Lateral; V, ventral. Scale bar: 100 μ m

3.2.4 Integration site and copy number of transgenes

X-Gal staining revealed that in most of the mouse lines the transgene expressing cells were distributed like OSNs expressing the endogenous mOR37C gene. Only in line #7 a slightly aberrant distribution pattern was observed. This raised the question whether in most cases the transgenic construct might have been inserted into the original gene cluster on chromosome 4 close to the endogenous expression control elements, whereas only in line #7 an integration into a novel genomic position might have occurred. To approach this question, FISH analyses on metaphase chromosomes were performed for lines #1, #3, #6 and #7, respectively. Double-FISH staining using a probe against BAC #422 M18 which spans a DNA segment adjacent to the mOR37 gene cluster (Strotmann et al., 1999) and a probe detecting the lacZ-part of the transgene revealed that in none of these lines the transgene had integrated into chromosome 4 (Figure 22). Using this technique, only one integration site for the transgene was observed in each line.

For six lines the distribution pattern of OSNs expressing the transgene thus very much resembled the endogenous counterpart, however, the number of cells was lower (Figure 16A-C) (see also table 2). In addition, significant differences in cell number were observed between the lines. The lowest number of cells was found in line #3, significantly higher numbers in lines #6 and #1; a remarkably high number of transgene expressing cells was found in line #7. It was next analyzed whether this difference in cell number might be due to different copy numbers of the transgene in the various mouse lines. Southern blots of genomic DNA probed with a lac-Z specific probe indeed revealed significant differences in the intensity of the reactive bands (Figure 21) indicating a considerable difference in transgene copy-numbers. To obtain a more quantitative estimate of this observation, real-time PCR experiments were performed. This approach revealed the highest number of transgene copies for line #1 (approximately 370) (table 2), whereas line #6 contained the lowest copy number (2-3). Line #3 which showed the lowest number of transgene expressing cells contained approximately 20 copies. Thus, there was no correlation between the number of transgene copies and number of X-Gal stained cells.

Mouse line	transgene copy number	Transgene expressing neurons	
		Heterozygous	Homozygous
Line1	370	158 ± 15	nd
Line2	60	---	nd
Line3	20	23 ± 3	nd
Line4	80	---	nd
Line5	85	---	nd
Line6	2	61 ± 8	114 ± 12
Line7	35	854 ± 75	~2000

Table 2 Number of transgene copy and the transgene expressing neurons in transgenic lines. nd: non-identified.

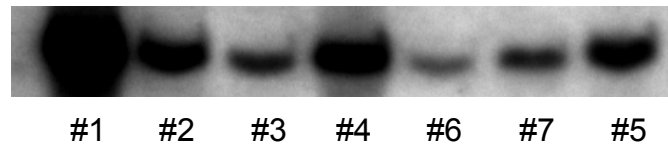


Figure 21 Detection of a 4758 bp fragment using probe specific for Lac-Z by southern blot. The genomic DNA from hemizygous mice was digested with EcoR I and a fragment of 4758 bp was detected with the Lac-Z specific probe. The mouse line is indicated beneath the southern blot lane.

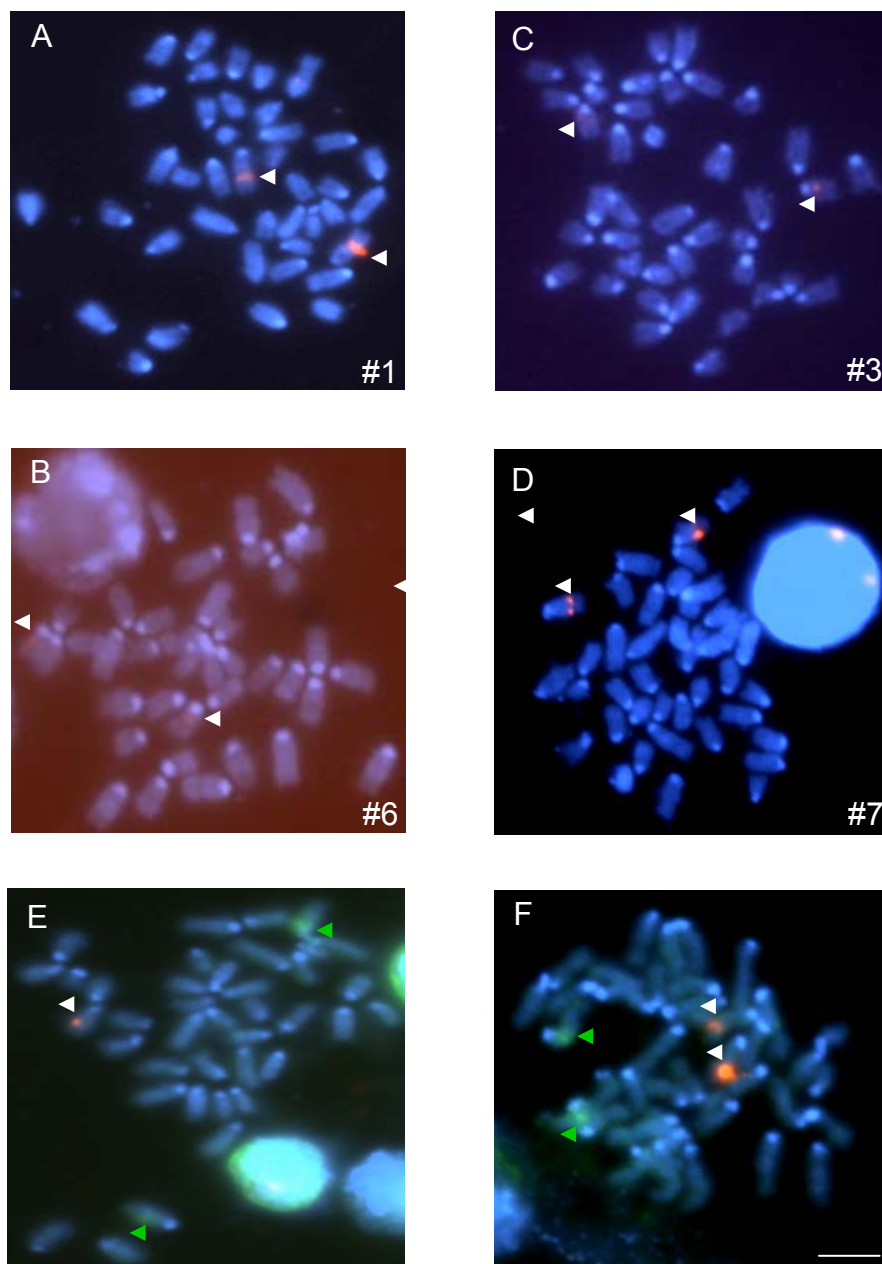


Figure 22 The integration site of the transgene. A, B, C and D show the integration sites of the transgene on chromosomes from Line1, 3, 6 and 7 homozygous mice. The chromosome

is DAPI stained (blue) and the red fluorescence indicated by the arrowheads mark the locus of the transgene. The size of the chromosome for lines 1, 3 and 6 indicates different chromosomes (E,F). Double in situ hybridization with probes detecting the *mOR37* locus on chromosome 4 and the transgene performed for line 7. Scale bar: 10 μ m.

3.2.5 Mutually exclusive and monoallelic expression of the *mOR37C* transgene

In order to determine whether the transgene and the endogenous gene were expressed in a mutually exclusive manner, transgenic lines were crossed with mouse line *mOR37C-GFP* in which the endogenous *mOR37C* gene is targeted by an IRES tau-GFP insertion such that cells which select this gene become green fluorescent (Strotmann et al., 2000). Inspection of large numbers (>200 in each line) of fluorescent OSNs in double-transgenic mice identified only cells with a single label (Figure 23A,B), indicating that the endogenous gene and the transgene were stringently expressed in mutually exclusive subsets of OSNs. By counting transgene expressing OSNs visible on the medial aspect of endoturbinates II in whole-mount specimen, it turned out that in hemizygous individuals (Figure 24A,B) only half as many labeled cells were visible as in homozygous animals (Figure 24C,D) (see also table 7); this is consistent with the idea that the transgenic alleles - like their endogenous counterparts (Strotmann et al., 2000; Vassalli et al., 2002) - were monoallelically expressed.

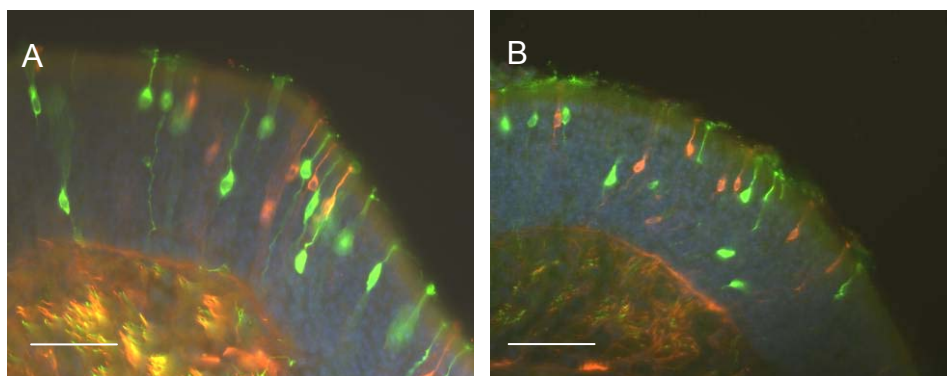


Figure 23 Mutually exclusive expression of the *mOR37C* transgene. Cross sections through the olfactory epithelium of double mutant mice hemizygous for the *mOR37C-lacZ* transgene (line #1) and heterozygous for the targeted *mOR37C-GFP* (A) or *mOR37A-GFP* (B). Cells expressing the targeted *mOR37C* gene were visualized by their intrinsic fluorescence, those expressing the *lacZ* transgene were detected by an antibody against β -galactosidase followed by an Alexa-568 labeled secondary antibody (red). No double labeled cells are detectable. Scale bar: 100 μ m.

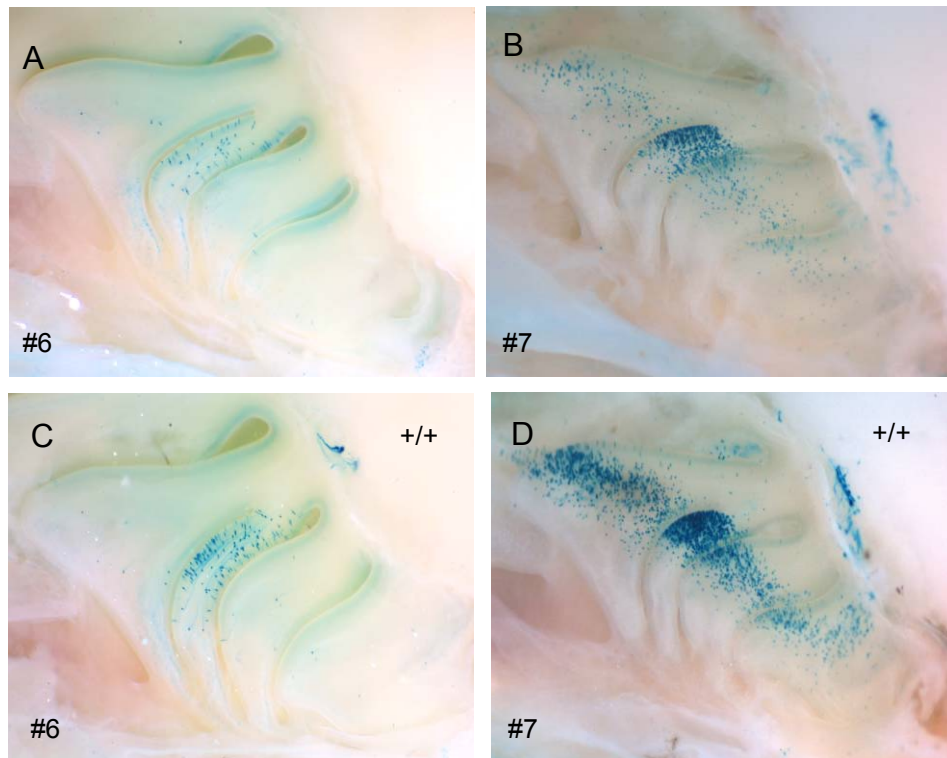


Figure 24 Monoallelic expression of *mOR37C-lacZ* transgenes. Whole-mount preparations of heads from hemizygous (A, line #6) (B, line #7) and homozygous mice (C, line #6) (D, line #7) mice showing the tau-lacZ marker revealed as a blue precipitate by incubation with Xgal (anterior to the right, dorsal to top).

3.2.6 Projection of transgene expressing OSNs to the olfactory bulb

We next analyzed the projection of transgene expressing cells in these lines by whole-mount X-Gal staining of the OB; axons became visible only after the bulb was carefully removed from the cranial cavity (Figure 25A-C). Stained fibers were then visualized on the anterior ventral surface which was closely attached to the cribriform plate. Axons converged onto a single position in the ventral domain of the OB which is reminiscent to OSNs expressing the endogenous gene (Figure 25D). In OBs from line #3 (Figure 25B) only faint staining was observed, consistent with the few OSNs expressing the transgene (see Figure 16B). Our whole-mount analyses thus suggested that the transgene expressing OSNs in lines #1 - #6 project to individual glomeruli, like their endogenous counterparts. This raised the question whether all the axons from OSNs expressing the endogenous gene and those expressing the transgene co-converged onto the same glomerulus. To address this question, double transgenic mice hemizygous for the lacZ-transgene and heterozygous for the GFP-

allele of the endogenous gene were analyzed. Serial sections through the OB were probed with anti-beta-galactosidase antibodies to visualize the axons of transgene expressing OSNs and anti-GFP antibodies for OSNs expressing the endogenous gene. In the majority of bulbs from line #1 the axons of OSNs expressing the transgene and the endogenous mOR37C indeed co-converged (Figures 26A-J); however, two patterns could be distinguished. The axons of transgene-expressing OSNs (red) either intermingled diffusely (50% n=14) (Figures 26A-C) with those expressing the endogenous gene, or a fraction of axons from transgene expressing OSNs formed an adjacent glomerulus (Figures 26D-F). In a few bulbs they formed a single glomerular structure which was split into a dense and less dense subcompartment (Figures 26G-J).

The projection pattern of transgene expressing OSNs in line #3 and line #6 followed very similar overall principles (Figures 26K-P). In general, the axons of OSNs expressing the transgene or expressing the endogenous gene co-converged and their axons mixed within a glomerulus. Subcompartmentalized glomeruli as seen in line #1 were not observed in lines #3 and #6. In rare cases, transgene-expressing OSNs did not co-converge with axons from endogenous OSNs.

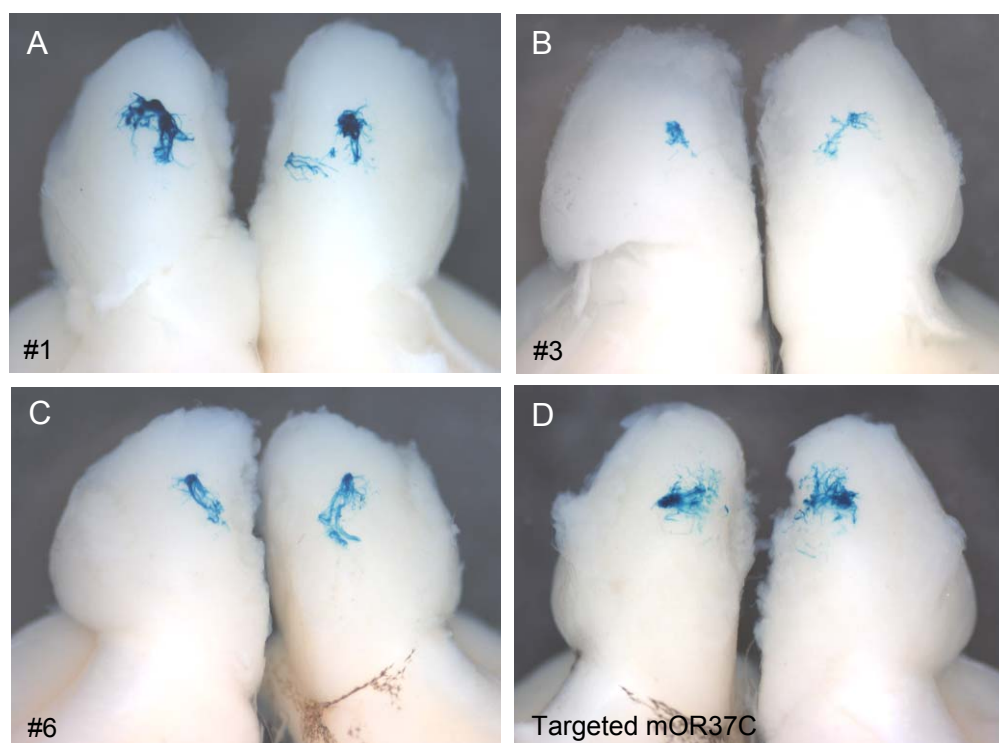


Figure 25 Whole mount view of the projection of neurons expressing the transgene. Bulbs were removed from the cranial cavity to expose the ventral surface. The tau-lacZ marker in axonal fibers was revealed by a blue precipitate after incubation with X-gal.

(A-C) In bulbs from lines #1, #3 and #6 axon bundles converging onto a single spot in the ventral domain are stained by X-gal; the projection pattern resembled that obtained for the targeted *mOR37C-lacZ* (D).

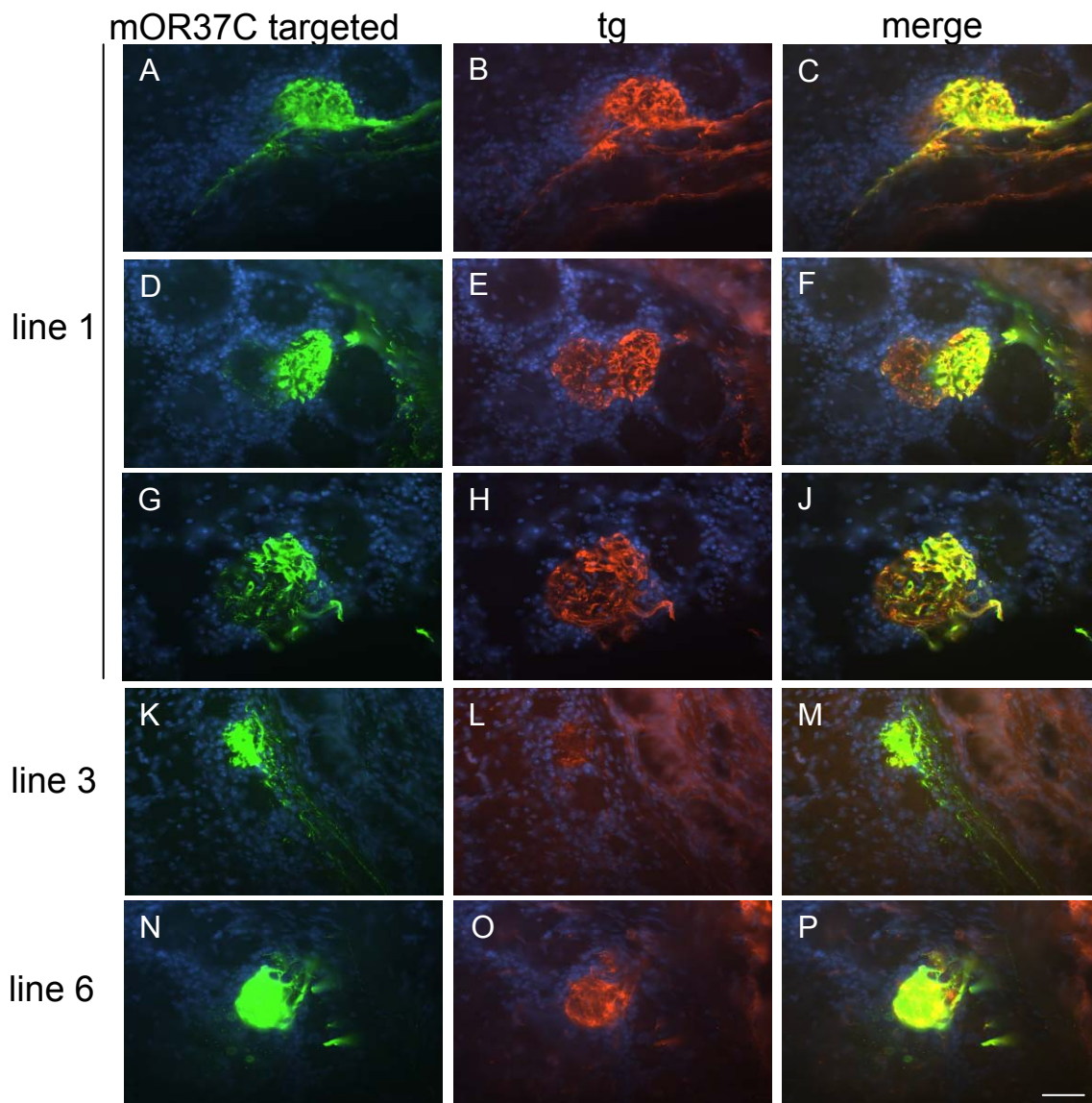


Figure 26 Projection pattern of axons from OSNs expressing *mOR37C*. Sections of bulbs from double mutant mice hemizygous for the *mOR37C-lacZ* transgene and heterozygous for the targeted *mOR37C-GFP*. (A;D;G;K;N) Axons from OSNs expressing the targeted *mOR262-12* gene visualized by their intrinsic fluorescence. (B,E,H,L,O) Axons from cells expressing the *lacZ*-marked transgene detected by an antibody against β -galactosidase followed by an Alexa-568 labeled secondary antibody. Axons generally coinnervate the same glomerulus in lines #1 (A-J), #3 (K-M) and #6 (N-P). In line #1, axons from transgenic *mOR37C-lacZ* cells also target neighboring glomeruli (D-F) or are segregated in subcompartments of the same glomerulus (G-J). Scale bar: 100 μ m in all pictures.

3.2.7 Effect of an ectopic *mOR37C* expression on the projection

The fact that in line #7 the transgene expressing OSNs were much more numerous than in other lines and that they were positioned outside the typical patch area raised the question about their projection. Of particular interest was the projection of OSNs which ectopically express *mOR262-12*; in view of the question of how much the axon targeting is affected by the position of cells versus the receptor-type they express. Whole-mount preparations allowing a view onto the medial aspect of the OB (Figure 27A) revealed blue staining close to the cribriform plate; the remaining medial surface of the bulb was devoid of staining. A view onto the ventral domain revealed thick fiber bundles which converged at several positions; the number of convergence sites ranged between 4-6 in each OB (Figure 27B,C). Thus, although the transgene expressing OSNs were widely distributed throughout the medial zone of the epithelium, their axons nevertheless all converged in the ventral domain of the OB. It could not be resolved whether these labeled structures indeed represented individual glomeruli. We next analyzed whether these axons co-converged with those from OSNs expressing the endogenous gene. On cross sections through the OB of double mutant mice the glomeruli formed by fibers from the endogenous OSNs also contained fibers from transgene-expressing cells, indicating that these two populations co-converged (Figures 28A-C). The glomeruli formed by these axons were larger than the glomeruli normally formed by the endogenous OSNs. As also seen in the whole-mount specimen, additional glomeruli were present; these were formed by axons from the transgene-expressing OSNs (Figures 28D-I). They were also located in the ventral domain of the OB, however, not positioned immediately adjacent to those formed by the endogenous cells (Figures 28D-F). In several of these transgene-glomeruli, a few GFP positive fibers were also detectable (Figures 28G-I), indicating the presence of axons from the endogenous OSNs in these structures.

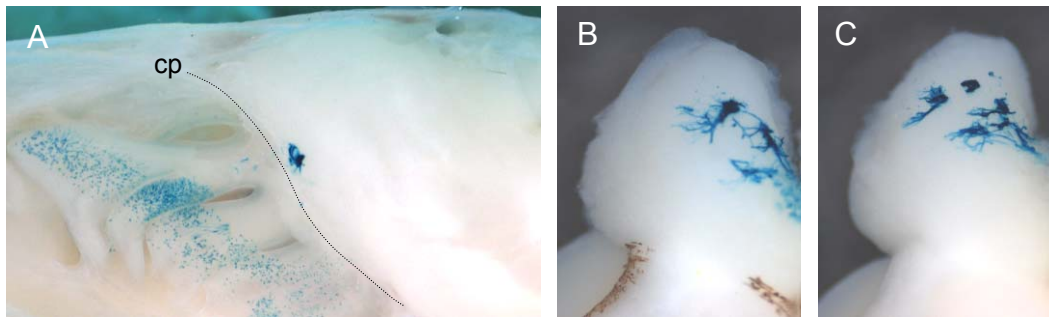


Figure 27 Projection of the neurons expressing the transgene in line7. (A) Whole-mount preparation of a sagittally transected mouse head, showing the medial aspect of the turbinates and the medial face of the olfactory bulb (anterior to the left, dorsal to top). The tau-lacZ marker was revealed as a blue precipitate by incubation with Xgal. Labeling in the bulb is visible only close to the cribriform plate (cp; dotted line). (B,C) Whole-mount preparation of bulbs removed from the cranial cavity; the ventral surface is exposed. The tau-lacZ marker in axonal fibers was revealed by a blue precipitate after incubation with Xgal. Several labeled spots on the ventral surface are visible.

In order to analyze whether the axons from transgene expressing OSNs in line #7 mistarget to glomeruli which express related ORs, double mutant mice hemizygous for the transgene and heterozygous for the GFP allele of *mOR262-14* (*'mOR37A'*; *olfr155*) (Strotmann et al., 2000) were analyzed. Although the axons of transgene expressing cells targeted glomeruli in immediate neighborhood of *mOR262-14* expressing cells (Figure 28J,K) these two axon populations were clearly segregated from each other; no axons from the transgene expressing OSNs were found in the wrong glomerulus. Together, the analyses demonstrated that the expression of *mOR262-12* in many cells of line #7 mice led to the formation of a few extra glomeruli which were positioned in the ventral domain of the OB, despite the fact that many of the neurons were ectopically located throughout the anterior-posterior extension of the epithelium.

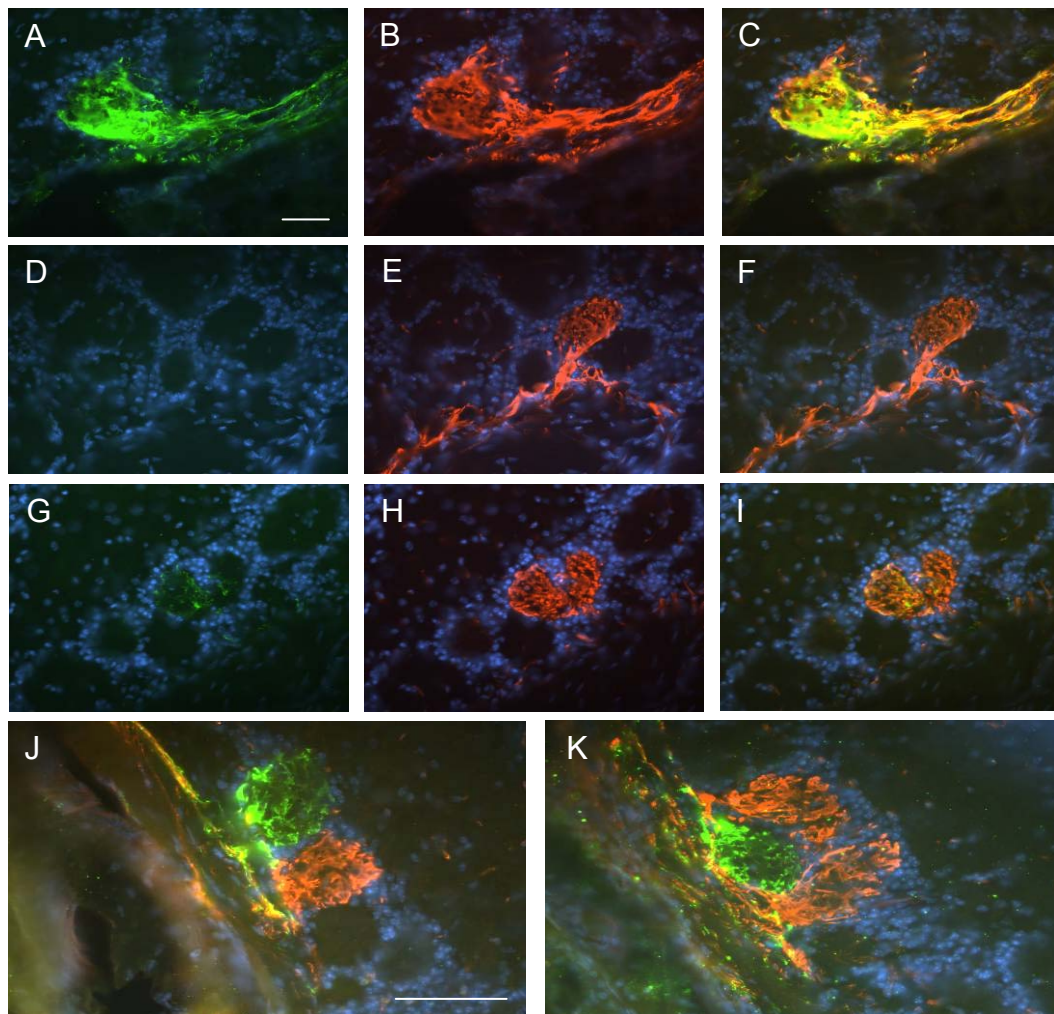


Figure 28 Projection pattern of axons from OSNs expressing *mOR37C-lacZ* in line #7. (A-I) Cross sections through bulbs from double mutant mice hemizygous for the *mOR37C-lacZ* transgene (line #7) and heterozygous for the targeted *mOR37C-GFP*. (A,D,G) Axons from OSNs expressing the targeted *mOR37C* gene visualized by their intrinsic fluorescence. (B,E,H) Axons from cells expressing the *lacZ*-marked transgene detected by an antibody against β -galactosidase followed by an Alexa-568 labeled secondary antibody. (A-C) Axon subpopulation from transgene expressing OSNs coinnervating the same glomerulus as those expressing the GFP-targeted *mOR37C*. (D-F) Axon subpopulation from transgene expressing OSNs forming novel glomeruli. (G-I) Axon subpopulations from transgene expressing OSNs forming novel glomeruli in some cases contain fibers from OSNs expressing the GFP-targeted *mOR262-12*. Scale bar: 100 μ m. (J,K) Cross sections through bulbs from double mutant mice hemizygous for the *mOR37C-lacZ* transgene (line #7) and heterozygous for the targeted *mOR37 -GFP*. Axons from OSNs expressing the targeted *mOR37C* gene visualized by their intrinsic fluorescence. (B,E,H). Axons from the two populations expressing highly related odorant receptors project to neighboring glomeruli; the axons are strictly segregated in distinct glomeruli. Scale bar: 100 μ m.

3.2.8 Search for the integration site of the transgene in line7

When homozygous mice were generated for investigating whether the transgene was expressed in monoallelic manner (see chapter 3.2.5), all lines behaved normally except line7; these mice constantly raised their head and cycled clockwise. Due to the atypical expression pattern of the transgene in line7 (Figure 16 and 28), this phenomenon was first assumed to be correlated with abnormal chemosensation. However, the normal behavior of hemizygous mice suggested that this phenomenon rather resulted from the insertion of the transgene to the novel genomic locus, thereby probably affecting the function of a gene residing at that position. Since in line 7 mice the vestibular system seemed to be affected, their auditory system was checked, as well. It turned out that homozygous mice also had a profound hearing deficit (Prof. M. Knipper, HNO Klinik, Tübingen). To get an insight into the underlying gene defect, a search for the integration site of the transgene was initiated using various strategies, such as inverse PCR, gene walking and adaptor mediated PCR (Figure 29). Using these techniques, it was tried to identify the transgene-genome transition. In each independent approach, a transition from one transgenic construct into another transgenic construct was found, confirming the RT-PCR experiments (see Chapter 3.2.4) which already indicated that around 35 copies of the transgene construct had inserted into the genome. In some cases, a transition from one construct into positions within neighboring construct were found (see figure 1), revealing that rearrangements of the transgenic construct had occurred in several of these copies during the insertion process; a phenomenon frequently observed for transgene integration events. However, from more than 200 PCR products no transition into a genomic region representing the transgene integration site could be found using these approaches. This seems to be resultant from the existence of multi copies of transgene which obviously favors the amplification of the segment from the tandem or from the rearrangement but prevents the fragment within the integration region from being amplified.

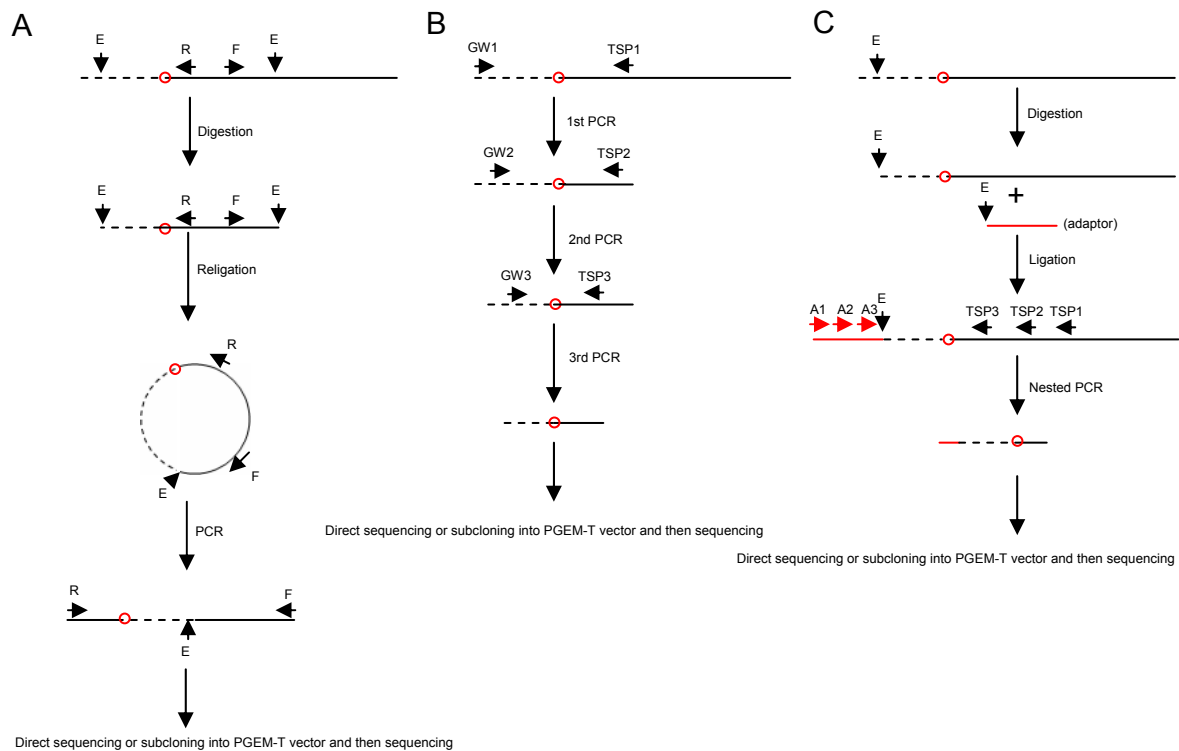


Figure 29 The diagram of the principle for Inverse PCR, Genewalking and Adaptor Mediated PCR.

A. The principle for Inverse PCR. The solid line indicates the transgene construct and the dashed line indicates the unknown flanking sequence. The red cycle indicates the integration site.

B. The principle for Genewalking. Nested PCR was performed with the primers from the transgene construct and from the Genewalking kit designed for the unknown flanking region.

C. the principle for Adaptor mediated PCR. The unknown flanking region was restricted and ligated with an adaptor (indicated by the red line) and the nested PCR was subsequently carried out with the primers from the transgene construct and from the adaptor. E: restriction enzyme; R: reverse primer; F: forward primer; GW: genewalking; TSP: target specific primer;

A: adaptor.

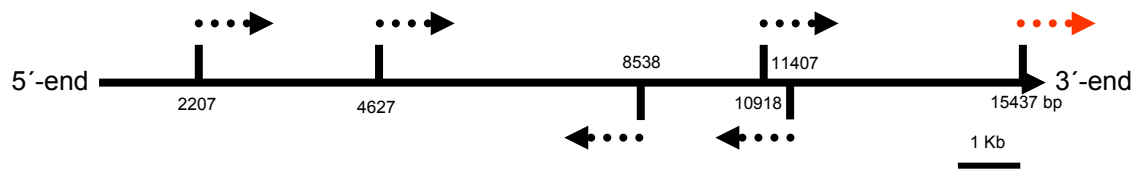


Figure 30 Overview of the rearrangements of the mOR37C transgenes. The black solid arrow shows one copy of the transgene and its orientation. The integration sites within the transgene are indicated by the vertical lines and the positions shown by the nucleotide numbers. The dashed arrows indicate other copies of the transgene and their orientation are the same as mentioned above. The red dashed arrow particularly presents the head to tail arrangement of the transgene and the black arrow indicates the rearrangement within the transgene construct. It should be noted that it is not necessary that all rearrangements occur within the same copy of the transgene since there are about 70 copies (see Table 2) of the transgene in the genome of homozygous Line 7 mouse.

3.3 No detection of interaction between H element and mOR37 promoter by Chromosome Conformation Capture (3C)

3.3.1 The principle of 3C

Chromosome Conformation Capture (3C) was first described by Dekker et al. (2002) and utilized to build up the scheme of chromosome conformation within the intact cell. It is based on the principle that subregions of the chromosomes which are physically located close to each other would be preferably cross-linked by DNA bound proteins upon treatment with a fixative (Figure 31A). This principle would imply that two segments which functionally interact, such as a promoter and a distal enhancer form a complex which is fixed upon cross-linking; in this context it would not matter on which chromosome or in which region of a certain chromosome the interacting segments are located (Ptashne, 1986). This procedure has previously been applied to investigation of how the transcriptional complex in the beta-globin locus is organised (Tolhuis et al., 2002) and also to explore whether the H element may interact with promoters of OR genes outside the mOR28 cluster (Lomvardas et al., 2006). To explore the possibility that the H element may interact with genes of the OR37 gene cluster, the chromosome conformation capture technique was employed to study exemplarily the gene mOR37C (31B).

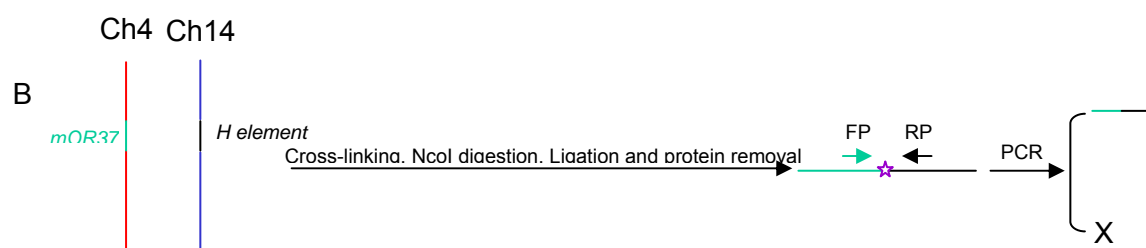
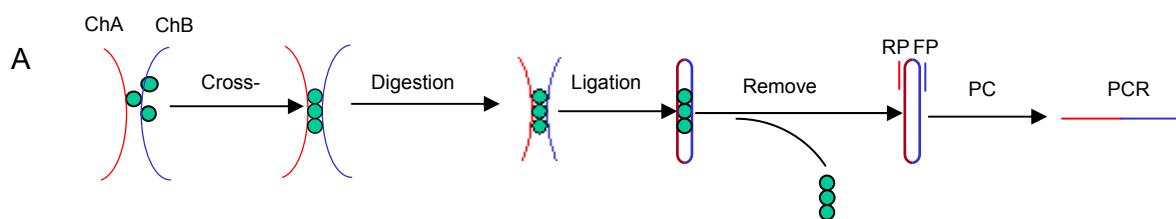


Figure 31 Diagram of the Chromosome Conformation Capture principle and the corresponding method to investigate the relationship of H element and mOR37C promoter.

A. The Principle of 3C. The green dots indicate the DNA bound protein. Ch: chromosome; RP: reverse primer; FP: forward primer

B. The procedure to detect the relationship between H element and mOR37 promoter.

✧ indicates the ligating site and X indicates there is no PCR product.

3.3.2 Searching for an interaction between H element and mOR37C promoter

According to the procedure of the “Chromosome Conformation Capture” chromatin was collected from OE or liver and cross-linked; the material was digested with NcoI and subsequently re-ligated. Primer pairs for the H element and mOR37C promoter were used for amplification. Figure 32A shows the result; only a very weak band was amplified from liver and from unfixed material of the OE, whereas a strong band was amplified from fixed material of the OE. The results indicate that this experiment apparently worked; the two controls were negative, neither from liver chromatin nor from unfixed OE-chromatin was a significant amplification product obtained. For further analysis the amplified band from fixed OE-chromatin was purified and subjected to sequencing. As expected two different sequences were obtained; as predicted one representing the mOR37C promoter, however, the other sequence did not represent the H-region but rather a sequence from chromosome 5 rather than from chromosome 14 where the H-region is located (Figure 32B and 33A). Comparing the region immediately upstream of the identified sequence on chromosome 5 with the primer used for H element revealed a striking sequence similarity; there were only 4 different bases, moreover 5 bases at the 3' end completely matched with the primer (Figure 33B). This feature explains the generation of an amplification product. These data do not support the notion for an interaction between the H element and the promoter but suggest that an element on chromosome 5 was interacting with the mOR37C promoter.

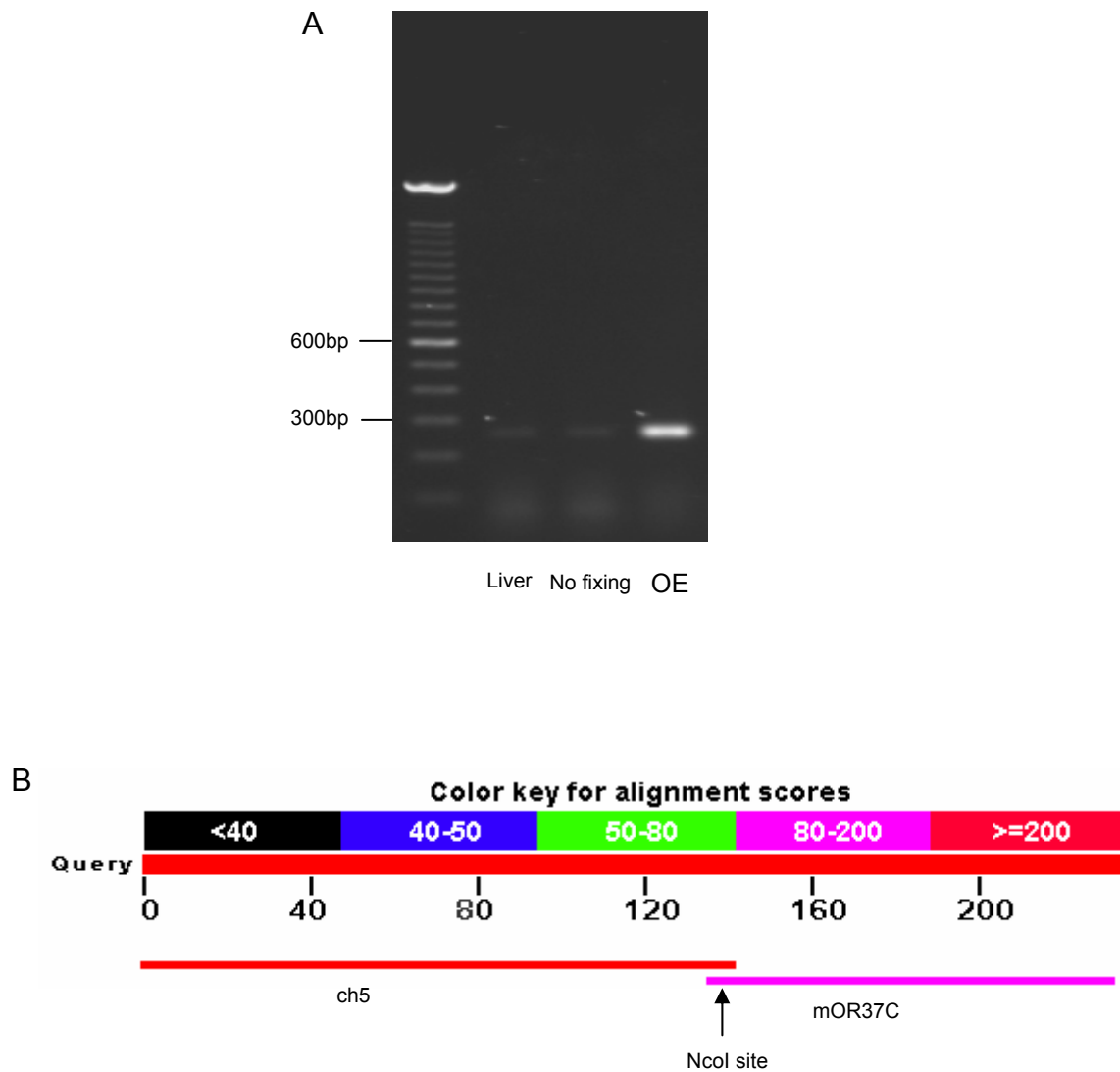


Figure 32 Identification of ch5 and mOR37C promoter interaction by 3C

A. Liver and OE without fixing generated little PCR product but OE produced extensive amplification.

B. Blast of the sequence of this PCR product against mouse genome found one part of the chimeric matched with mOR37C promoter and the other matched with ch5 sequence. NcoI site was found between these two parts.

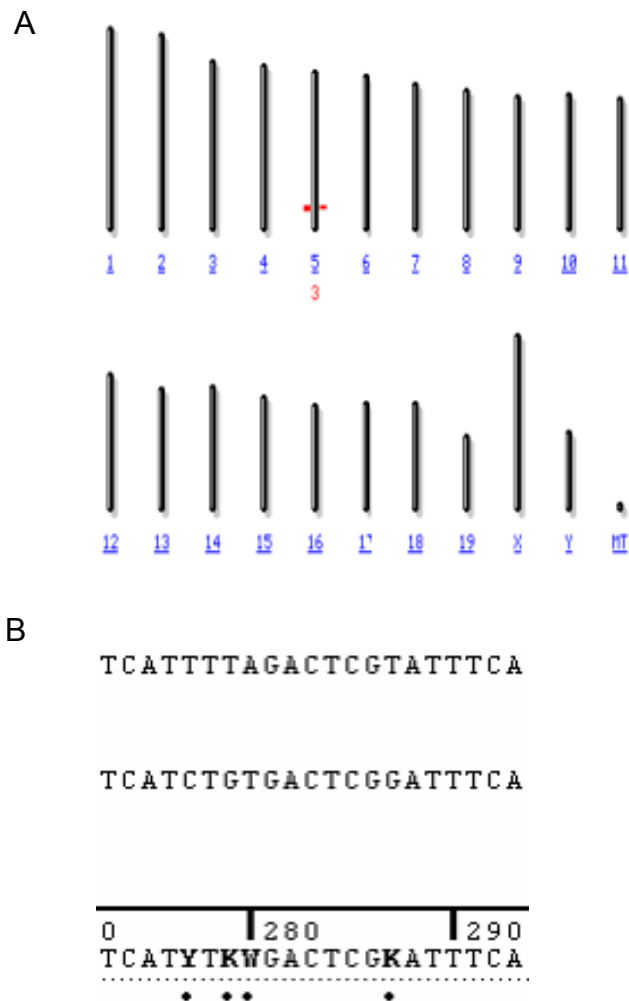


Figure 33 Comparison of H element primer with the upstream sequence of the ch5 region.

A. The location of the sequence on ch5. This region was downloaded to compare with the primer for the H element.

B. Comparison between the H element primer and the immediate upstream sequence on chromosome 5. A clear sequence similarity emerges; there are only 4 different bases, moreover 5 bases at 3' end are completely identical.

>NcoI fragment including the 153 base pairs from 3C.

```

ccatggccatagctttgttttagggtagttctttatcttctgtctttggctttatcgattgtttatatattgatcctctaacttaatta
gaatggtagcccttatcactataatccctttacagctggttgttttacctttgtgaaatacgagtctaaaatgacattcttcta
atagaaagtcaaaggaagggaaaagaaaataaagaaaagaagcatagaccgaaaaggaatctggcctaagttc
agtttctctgtctgaatctcaagttctggttttaaaactgaaatgctgtagactcatctgtacttgcaagacatagtgaga
aggagcctcgtctactttgcctgatctccttaatgctatctgtgttttctcttctgttgccaataaaactccaatgaag
ggaatgaggcctaactttggctctctgtccaagagaacacagtccatcatggcaggaggcagcctggcagcaggagt

```


gtaaggcagctggttacactgtatcaagaagcagagagagaggctgatggctatagtcagagaagggaaaattata
ctcctttgggaagaaggccactgaaaggtgcttatggctcaggggatgatcatcacatacatgcatgtgggaagctct
cagtggacttgggtcatcagtaagacaaatagagaggacaggaagctgggaaggagaagaaggatgccaggg
agggttggtggtgacagtgtatagtcacgggtgggagacaagaggaaaacaactggcactgacttctggcctctaca
caggcacacatccaatTTTTTTTTTgaagaattTGGCAAagagtctagtatgacatagggtgtttatacaaggatgaca
tacggaacccttgctcatggcgatgaaatgacttgggtgacttgagtttgatgtcttcattagctctctttgcccctccccct
TTTTGagacagactcttTcactggacctagagctcagcaagtaggcaagaatagctggccagtgagccccagtgatc
ctctgtctctgctgcacagtgtactaatacaggctcgcacaccacacctggactatttatgtaggaggtggaggtctg
aatttaagccctcatgcctaagtgacaagcccttTcctaattgactgagccatctcccctggcccatgatgtacacttctt
aactttatataaggataactggaaattttTtacacatac
agatcctaaagccacctttactggt
atatcacctgctcagtagtattgtgtgctgagtatgcaccgtggttttatTaccattcactgttgttatatagctgagctgtt
acatttctagctgtgatgaataaagctgctgtcaatagttataagcaggtttctatgaataatgcattttcacattttttatTta
attttattctttgggtaaagtatctggagtgcaattgcagggtactatggtgattccatgttccattttgtgacaaaatacca
agtttccctgctgtgctggcaacattttataTTTtataaatcaatgaataaatttctctccctcttccctctccctctcacc
ccttccccttccctctcttccctcttctcttctgtgtgtgtatgtgtatacatggatttagttaccctcagaagccagaaggct
gtgttgattctctgaagctagagttataggcaactctgagctgctcagtgtaggtgctgggatctaaacttgggtcctttgt
aggagcagaatataactttTaa**ccatgg**

Figure 34 Sequence of the ch5 NcoI fragment interacting with mOR37C promoter by 3C. The NcoI restriction sites are indicated in green. The nucleotides indicated in red are the ones obtained directly from sequencing of the 3C PCR product.

3.4 Search for locus control region-like elements of the mOR37 cluster

3.4.1 Background

In mammals, most of the chromosomes comprise OR genes, often at many loci; for instance on 21 human chromosomes there are 51 loci, and 51 loci are also on 17 mouse chromosomes; highly related ORs are often clustered in one locus (Malnic et al., 2004; Godfrey, 2004; Zhang et al., 2002). The mechanisms controlling the expression of genes which are arranged in one cluster have been extensively studied for the beta-globin locus. For the clustered globin genes it was shown that the

selective expression of each member is controlled not only by the promoter but also by a distally positioned locus control region (LCR) (Grosveld et al., 1987; Fraser et al., 1998). The operation of a LCR is supposed to be achieved by opening chromatin at the locus and then forming a complex of the LCR with the promoter of a particular globin gene by a long range looping (Carter et al., 2002; Tolhuis et al., 2002; Li et al., 2002; Bulger et al., 1999). The number of gene clusters which are under the control of such LCRs or LCR-like elements is gradually increasing based on the results of both computational and experimental investigations (Li et al., 1991 and 1999; Li et al., 2002; Bulger et al., 1999). In view of the mechanisms active in controlling the globin genes the question raised whether the clustered OR genes may also be under the control of an LCR-like element. By comparing candidate genome sequences from human and mouse, an LCR-like segment, the H element, was identified as a control element for the mOR28 gene cluster (Serizawa et al., 2003); subsequently it was suggested that the H element may be important for many, if not all OR genes (Lomvardas et al., 2006). However, all attempts to demonstrate any interaction of the H-region with mOR37 genes failed (Chapter 3.3). Since the OR37 genes show a unique expression pattern (Kubick et al., 1997; Strotmann et al., 1999) it seems possible that the mOR37 cluster may have a unique LCR for regulating the expression of these genes. Attempts to identify LCR-like sequences which may be relevant for the mOR37 cluster were performed.

3.4.2 Unavailability of identifying mOR37 cluster I related LCR like elements by sequence comparison across closely related species

The sequences of the syntenic conserved OR37 cluster I from the human, mouse and dog genome (Hoppe et al., 2006b) were downloaded and an alignment was performed with the program *PipMaker* (Schwartz et al., 2000). Given the high conservation of the flanking genes, the cluster-region was limited to the immediate non-OR gene on both sides. This could miss the LCR like elements if they were located far distant from the locus, however, this approach could increase the probability that any conserved elements to be identified may indeed belong to the OR37 cluster rather than to other genes. Using the mouse sequence as the reference, the *PipMaker* aligned the three sequences as shown in Figure 35A and the local similarity is shown in Figure 35B. A few segments which are conserved among all three species were identified. The sequence examination from the Text Data indicated that the sequences with highest similarity were the coding exons of

the OR genes (shown as red column in A and red arrows in B). Therefore, the other regions were taken into consideration. The three segments enclosed by blue squares, one upstream, one within and one downstream of mOR37 cluster I, were in the non-coding region of the OR genes, indicative of potential candidates for LCRs. To exclude the possibility that they may be coding exons of other, possibly non-OR genes, the sequences from these three segments were downloaded and blasted against the mouse genome. The first upstream as well as the downstream segment did not hit any protein coding sequence; parts of the second upstream segment identified two genes predicted genes by NCBI (Figure 35C). In addition, some short segments close to the coding regions of the OR37 genes were also found to be conserved among all the species, these regions were defined as non-coding exons or promoters of the mOR37 genes (Hoppe et al., 2003b). Furthermore, some regions were found which were only conserved between two species. Due to the syntenic conservation for this locus across all the species, it was not possible to determine a candidate LCR-element by computational investigation.

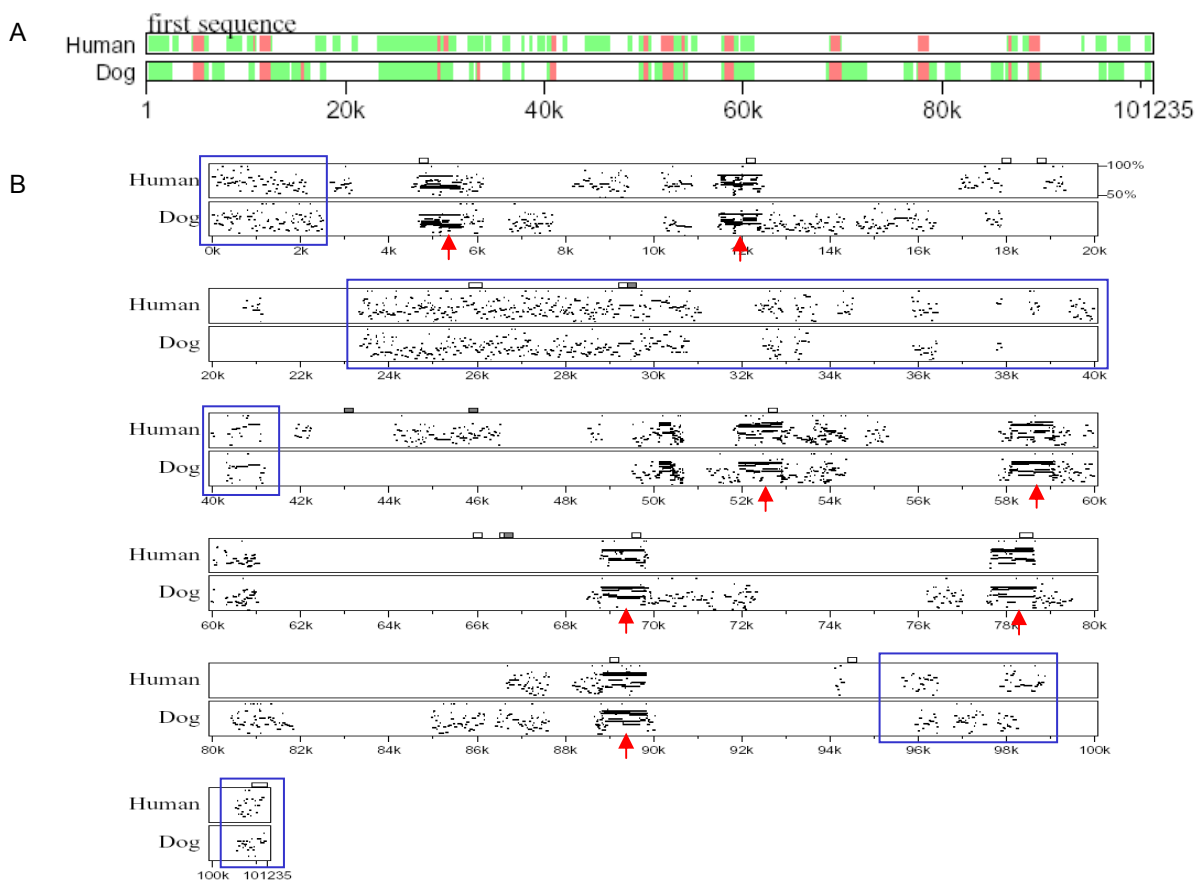




Figure 35 Detection of conserved segments in the OR37 cluster locus among human, mouse and dog genome.

A. With mouse cluster I as the reference sequence multi alignments was performed by Pipmaker. The basal identity for showing the alignment is 50% and the high and low similarity are presented by the red and green columns, respectively.

B. The detailed segments of similarity from A. The blue squares enclosed three big segments, two upstream and one downstream of the cluster I, conserved among the three species. The red arrows indicate coding region of the five mOR37 genes and two evolutionary related OR genes olfr70 and olfr71.

C. The gene annotation of the region between mOR37E (olfr159) and olfr71. The red arrows indicate the protein coding sequences predicted by NCBI.

The availability of genomic sequence data from ancient mammals, like Opossum and Platypus, now allows a more extensive pair wise sequence comparison and in addition a multi-species sequence comparison: these approaches have previously been employed to identify LCR of other genes (Regales et al., 2003; Li et al., 1991 and 1999).

3.4.2.1 Conservation of mOR28 cluster across multi-species

In order to validate the applicability of this comparative approach on sequences from very distantly related species, including Opossum and Platypus, attempts were made to determine whether the H element can be found across the species borders. As a first step it was determined if genes of the OR28 cluster are conserved among the investigated species. For this goal sequences of the OR28 cluster from mouse, such as mOR28, mOR10, mOR83 were used to perform blast searches in the genome of other species to identify homologous genes. The blast results together with the annotation by means of the Ensemble database (<http://www.ensembl.org/index.html>) revealed that this cluster was in fact conserved across the species, including mouse, human and dog, but also opossum and platypus (Figure 36 and Table 3). For all species, except dog, the loci showed a syntenic conservation; i.e. also the neighboring genes of the loci, including the T cell receptor gene cluster, were conserved.

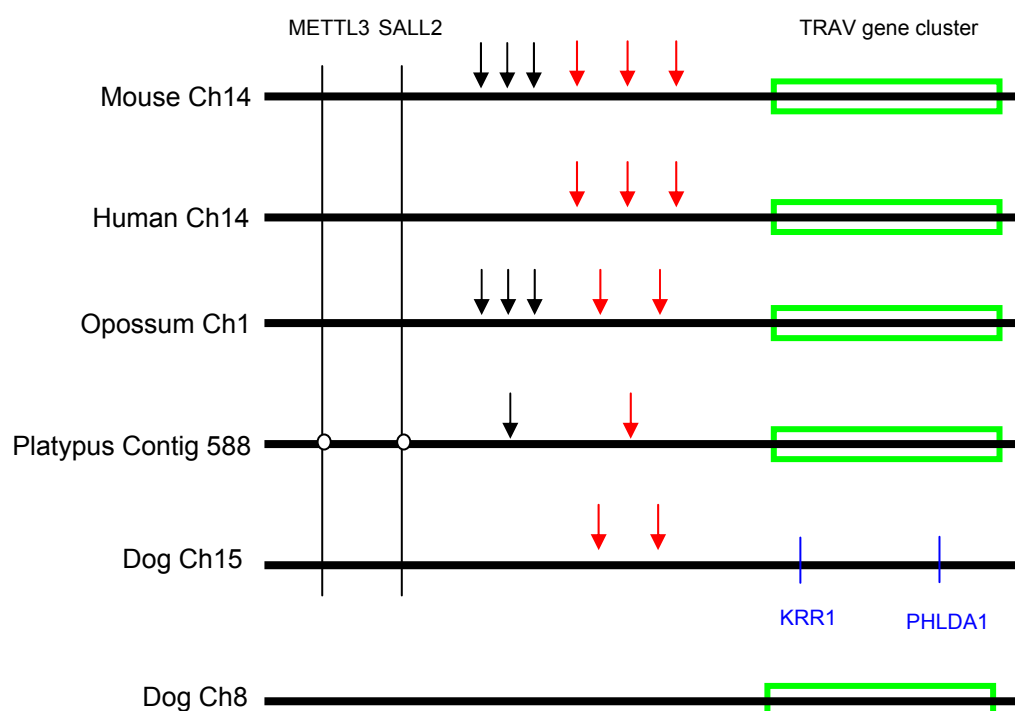


Figure 36 The conservation of OR28 loci in mouse, human, dog, opossum and platypus. The olfactory receptor genes are indicated by arrows, of which the red are the orthologs of mOR28, mOR83 and mOR10 (subfamily mOR244) and the white are the olfactory receptor genes belonging to subfamily mOR223. The syntenic conservation of the locus is defined by the flanking genes. The locus upstream genes are highly conserved in all species (indicated by the vertical lines) except for platypus (the assembly has not been finished for this region indicated by the white dots). The downstream TRAV gene clusters are shown by the green squares for all species except dog for which the downstream genes are variable and shown by blue vertical lines. The TRAV receptor gene cluster is positioned on chromosome 8 instead of chromosome 15 in the dog genome. Ch: chromosome.

Species	Genomic location	OR gene number
Mouse	Ch14: 51250713--51489244	6
Human	Ch14: 21075177--21250394	3
Dog	Ch15: 21578389--21768579	2
Opossum	Ch1: 170150142--170467178	5
Platypus	Contig588: 486031--514993	2

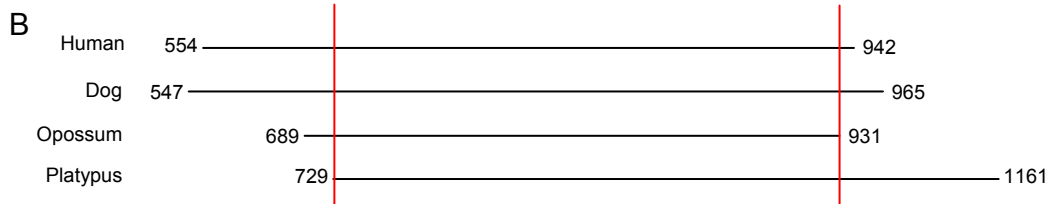
Table 3 The position and the olfactory receptor gene number of mOR28 locus across species.

3.4.2.2 Does sequence comparison allow to identify H element sequences in ancient species?

Based on the existence of the OR28 cluster in all species, analyses were performed trying to determine if an H-region can also be found in all species. The H element sequence of mouse (Serizawa et al., 2003) was used for the subsequent blast analysis. In all four species the best hits are always targeted to a region of the OR28 cluster. The sequence of the putative H element from various species was subsequently compared to evaluate if any particularly conserved sequence motifs may exist. Comparison between mouse and human revealed four conserved segments but between mouse and dog only three segments were conserved; these three parts were also conserved between human and dog. When opossum and platypus were taken into account, only one conserved segment was found, but interestingly this segment was conserved among all the species (Figure 37A and B). This sequence was determined to comprise 202 bases corresponding to the region between position 729 to 931 in the mouse H element sequence (Figure 37B and C). This motif may be considered as a core sequence for the H element.

A

	Links	Query			Chromosome			Chromosome			Stats					
		Start	End	Ori	Name	Start	End	Ori	Name	Start	End	Ori	Score	E-val	%ID	Length
Human	[A] [S] [G] [C]	<u>554</u>	942	+	Chr:14	21241796	21242180	+	Chr:14	21241796	21242180	+	166	3.7e-115	72.47	396
	[A] [S] [G] [C]	1306	1790	+	Chr:14	21242589	21243080	+	Chr:14	21242589	21243080	+	160	3.7e-115	67.66	504
	[A] [S] [G] [C]	1043	1306	+	Chr:14	21242307	21242570	+	Chr:14	21242307	21242570	+	129	3.7e-115	75.28	271
	[A] [S] [G] [C]	216	519	+	Chr:14	21241432	21241731	+	Chr:14	21241432	21241731	+	52	3.7e-115	61.25	320
Dog	[A] [S] [G] [C]	547	965	-	Chr:15	21681331	21681744	-					156	2.4e-69	70.16	429
	[A] [S] [G] [C]	1344	1785	-	Chr:15	21682210	21682644	-					133	2.4e-69	67.11	456
	[A] [S] [G] [C]	1042	1130	-	Chr:15	21681852	21681938	-					45	2.4e-69	76.34	93
Opossum	[A] [S] [G] [C]	<u>689</u>	931	-	Chr:1	170437408	170437655	-					59	2.4e-06	64.73	258
platypus	[A] [S] [G] [C]	729	1161	-	Contig588	492226	492647	+					43	0.068	58.55	456



C

```
>202bp from 729 to 931 in mouse H enhancer
agagttagttacttttgacaagttacagatccacagggattatctg
ataagatgaacaaccagaggaaaacataactttttaatgagtcctc
ctgtgttttctggttcatgctaacatcactggtaagacatagggt
gcctctagtggttcattatagctcctctctcattaattgccctctg
ctctcaaaatgcattctct
```

Figure 37A Overview of the blast using the mouse H element against other species. The region and the orientation of the query and identified sequence are shown by the base number and +/- . The statistic values such as Statistic score and E-val are also displayed followed by the sequence identity and sequence length. The green underline indicates the segment overlapping in all species.

B. The green line indicates segments which are aligned and the red vertical lines define the overlapping region.

C shows the sequence of the overlapping region in B.

3.4.2.3 Is the “H element” in dog related to OR28 cluster or TCR gene cluster?

Interestingly, in dog the T cell receptor gene cluster was missing from the vicinity of the OR28 locus, but other neighboring genes showed a syntenic conservation. Blasting the dog genome with the mouse T cell receptor genes revealed that in dog the conserved cluster was located on another chromosome (Figure 36). This unique arrangement of the conserved mOR28 cluster on chromosome15 and the T cell receptor gene cluster on another chromosome may allow to determine if the H element was associated with the OR genes or was segregated with T cell receptor genes.

Since the OR28 locus and the T cell receptor locus are separated in the dog genome, the T cell receptor locus was also carefully examined for an H element sequence, however, no indication for a H element was found in this locus (Figure 38). This result suggests that upon segregation of OR28 and T cell receptor clusters the H element was assigned into the OR28 cluster. The close association indicates the relationship between the H element and the OR28 cluster. Thus, the syntenic conservation with the OR28 cluster instead with the T cell receptor gene cluster supports the findings of Serizawa et al. (2003) and strongly implies that the H element is a specific component for controlling the expression of the OR28 cluster genes.

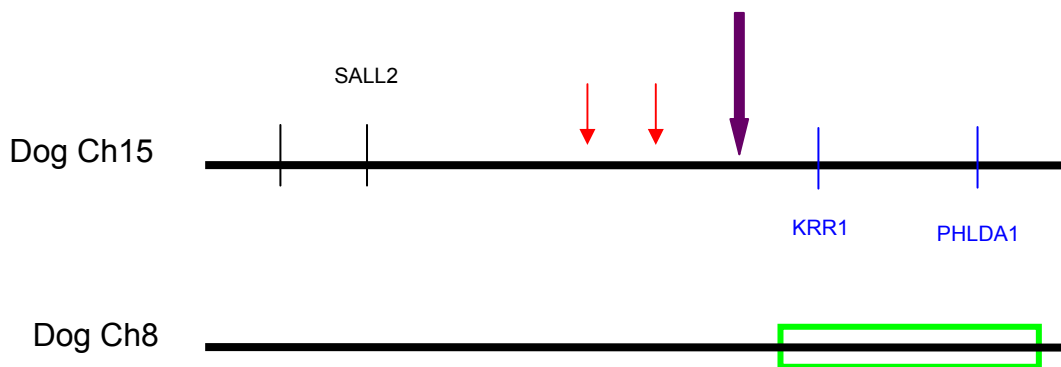


Figure 38 Colocalisation of the H element homology with the mOR28 cluster instead of the TRAV cluster in dog. The figure annotation is the same as Figure 36. The purple arrow block indicates the blast hit the mouse H element in dog genome.

3.4.2.4 Attempts to identify an H element sequence by means of multi sequence comparison

A conserved 202 bp segment of the H element was identified by a pairwise comparison between sequence from mouse and one other species (Figure 37). To confirm this result, attempts were made to apply a multi sequence comparison

approach (Schwartz et al., 2000). Towards this goal the sequence of the OR28-related locus from all five species was downloaded from the Ensemble Data base and all repetitive elements were removed. Subsequently, the multi sequence comparison approach was performed using the program *MultiPipMaker* (Schwartz et al., 2000); the mouse mOR28 locus was employed as the reference. Figure 39A and B show an overview of the alignment; in all species, four segments were found to be conserved (Figure 39A, arrows). Examination of the sequences revealed that three of them corresponded to conserved regions of the OR genes itself (indicated by the red arrow in 39A). Therefore, the subsequent investigation was focused on the fourth conserved segment (indicated by blue arrow in Figure 39A and displayed in more detail in Figure 39B by a blue arrow). A sequence alignment of the fourth segment from all 5 species is shown in Figure 40 documenting the high degree of sequence identity between the phylogenetically distant species. The 310 bases segment was then compared with both the H element reported by Serizawa et al., (2003) and the 202 bases segment identified as described in chapter 3.4.2.2; this comparison was performed using the program *ClustalW*. The result is shown in Figure 40B. The sequence of the H element is displayed in the middle; the sequence obtained by the pairwise comparison is positioned above and the sequence obtained by the multi sequence comparison is positioned below. Sequence identity is indicated by stars located in the fourth line. Figure 40B clearly demonstrates a complete identity of all three sequences between position 730 and 931 of the H element sequence reference. The data demonstrate that both approaches are able to identify an H element sequence not only in mouse and human (Serizawa et al., 2003) but also in very ancient mammalian species, like opossum and even platypus. In addition, both comparative approaches led to the identification of a particular 202 base pairs element which is highly conserved between all species studied ranging from platypus to human.

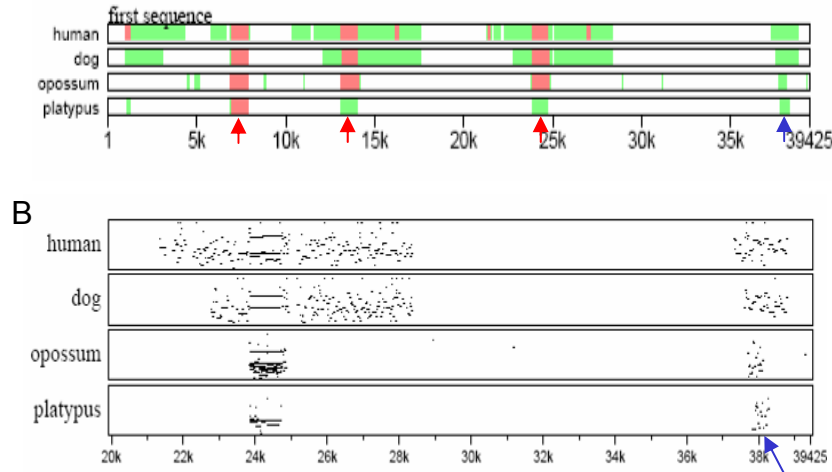


Figure 39 Multi sequence comparison of the OR28 locus by *PipMaker*. An overview of the alignment is shown in A and part of the identity overview is presented in B. The red arrows indicate the conservation of mOR10, mOR83 and mOR28 across the species. Additional non OR conservation is indicated by blue arrows in A and in B.

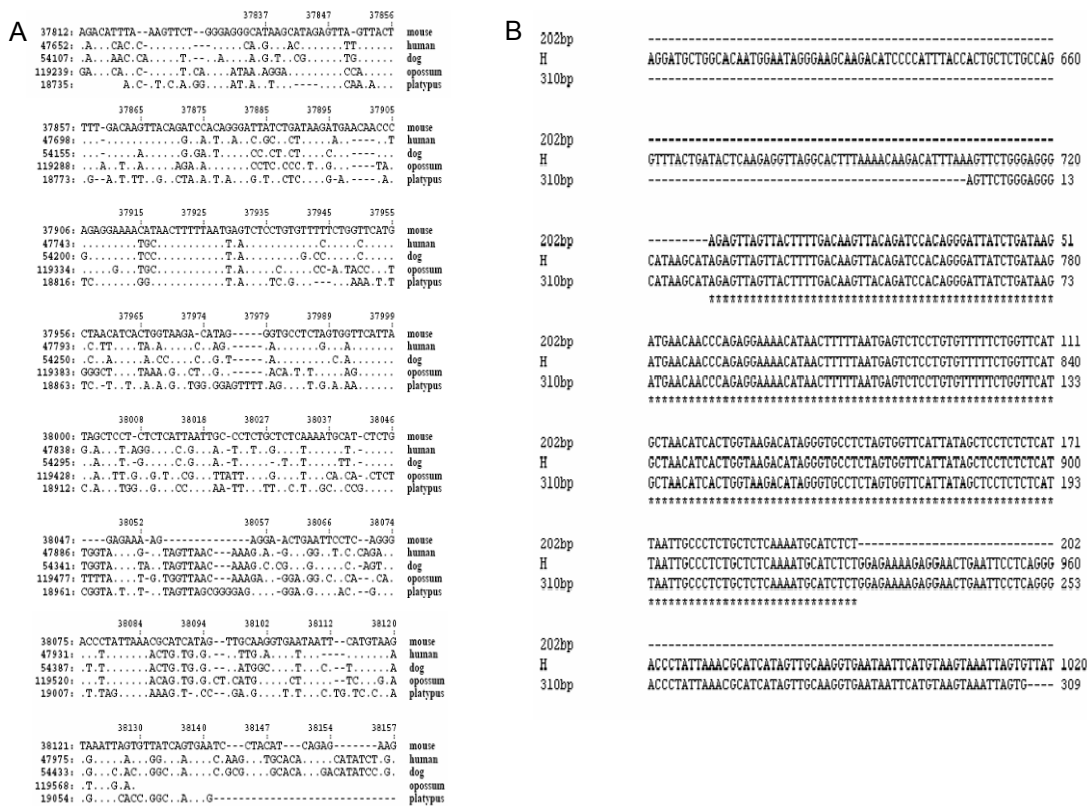


Figure 40 A. Details of the sequence alignment of all species from the region indicated by the blue arrow from 39B. Overview of the alignment of the corresponding mouse sequence (310 bases) from this alignment, the mouse H element and 202 bases of the segment identified by Blast in chapter 3.4.2.2.

3.4.3 Comparison of mOR37 gene clusters from different species

Since the multi-species approach led to the identification of the H element for the OR28 cluster in 5 mammalian species, now it seemed suitable to apply the same procedure to the OR37 cluster. As a prerequisite for this approach it was necessary to evaluate if the gene clusters and genomic loci really exist in all species under investigation. It has previously been shown by our group (Hoppe et al., 2003a; Hoppe et al., 2006b) that for the OR37 cluster II a conserved syntenic locus existed in both Eutherian and Metatherian species; whereas the OR37 cluster I was only found in the Eutherian species but not in Metatherian species; this was confirmed by a recent study (Aloni et al., 2006). However, when the previous studies were performed the genome sequence of opossum was only roughly assembled into scaffolds and the sequence quality as well as the gene annotation was still in a preliminary status. With the currently improved resolution of the sequence and the updated assembly it was possible to assess whether the opossum genome comprises in fact only OR37 genes corresponding to cluster II or if OR37 genes related to cluster I may also exist in opossum and platypus. Blast analyses with the coding sequence of mouse mOR37 cluster I genes were subsequently carried out with the two genome sequences. The best hits in the opossum genome targeted a gene cluster containing 12 OR genes (Figure 41 and Table 4 and 5). The best hit in the platypus genome targeted only one OR gene which was found to be homologous to mouse olfr270; confirming previous findings (Hoppe et al., 2006) and indicating that cluster I may not exist in platypus. The opossum genes were subsequently further examined by Ensemble annotation and by comparison with other species (Figure 41 and Table 4). Listing the opossum genes (Table 5) clearly showed that the composition of this cluster highly resembled that of mOR37 cluster I in other species. The finding that there are diverse copies of one gene as well as one gene related to cluster II suggests, that the identified group of genes may be a most ancient cluster I.

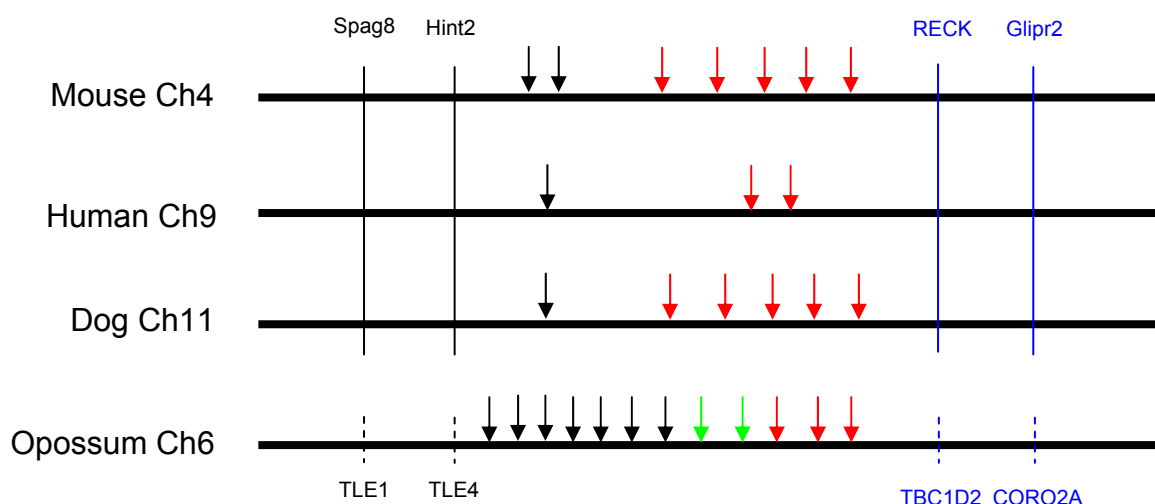


Figure 41 The conservation of mOR37 gene cluster I in mouse, human, dog, opossum. The olfactory receptor genes are indicated by the arrows, of which the red are the orthologs of mOR37 gene and the black are the OR37 related olfactory receptor genes belonging to subfamily mOR262. The green arrows indicate the OR37 gene similar to OR37 gene cluster II in opossum. The syntenic conservation of the locus is defined additionally by the flanking genes. The black and blue vertical solid lines indicate that the upstream and the downstream genes are conserved across species. The dashed vertical lines indicate the variable flanking genes in opossum genome. Ch: chromosome.

Species	Genomic location	OR gene number
Mouse	Ch4: 52633263--53010410	7
Human	Ch9: 105943519--106549790	8
Dog	Ch11: 63376121--63743921	7
Opossum	Ch6: 141275491--141996688	12
Platypus	Ultra477: 455081--550219	1

Table 4 The position and the olfactory receptor gene number of the OR37 cluster I locus across species.

Hoppe et al	Zhang and Firestein	NCBI
---	MOR262-10	Olf70
---	MOR262-4	Olf71
mOR37E	MOR262-5	Olf159
mOR37D	MOR262-11p	Olf158
mOR37B	MOR262-6	Olf156
mOR37C	MOR262-12	Olf157
mOR37A	MOR262-14	Olf155

Table 5 The composition of mOR37 cluster I and the distinct nomenclature of the corresponding OR genes according to Hoppe et al., 2003; Zhang and Firestein, 2002 and NCBI. The non-OR37 genes are indicated by the dashed lines. The genes are outlined in the order from centromere to telomere.

locus composition	Location	Homologous Annotation
Vps13D, Vps13D, 12 ORs, TBC1D2, Coro2	Ch6: bp 137679377--139061902	---
OR1	Ch6: bp 138432462—138446315	Olf71
OR2	Ch6: bp 138446446—138449975	Olf70
OR3	Ch6: bp 138470594—138471505	Olf70
OR4	Ch6: bp 138492426—138493355	Olf70

OR5	Ch6: bp 138516934—138517863	Olf70
OR6	Ch6: bp 138532683—138533612	Olf70
OR7	Ch6: bp 138549243—138550199	olf159 or Olf270
OR8	Ch6: bp 138560979— 138561953	Olf273
OR9	Ch6: bp 138585635—138586597	Olf272
OR10	Ch6 bp 138624444—138627966	Olf156
OR11	Ch6 bp 138646265—138648418	Olf157
OR12	Ch6 bp 138687533—138688492	Olf155

Table 6 Gene annotation of the novel Opossum OR37 cluster I by Ensemble. The locus is comprised of two upstream non OR genes, 12 OR genes and 2 downstream non OR genes. The order of OR genes is based on the gene position on the forward strand and the OR gene span predicted by Ensemble is shown by the starting and ending base pair position. Olf155, olfr156 and olfr157 correspond to mOR37A, mOR37B and mOR37C respectively. Olf272 and olfr273 are two OR37 genes from cluster II. Olf270, olfr70 and olfr71 are mOR37 related genes, namely from the same subfamily but without the typical six amino acid insertion in the third extracellular loop, from cluster II and cluster I. The genes are outlined in the order from centromere to telomere.

3.4.4 *Pipmaker* analysis of the cluster I locus plus opossum

Due to the syntenic conservation for this locus across all three species, it was not possible to determine a candidate LCR element by computational investigation (Figure 35). Therefore, cluster I from opossum was included in the subsequent analysis. Since the opossum cluster I provides the advantage that the flanking genes in the locus are not homologous to the ones in mouse, dog and human. Another benefit would be that the length of the locus sequence could be extended, which

appeared particularly important for distal region of the cluster. Figure 42A shows the alignment performed by *Pipmaker* using the mouse sequence as the reference; ten segments emerged which were conserved among all species. Close examination of these sequences revealed that eight of them were the coding sequences of OR genes, one was part of a coding sequence from a non OR gene (left purple arrow) and one was the promoter of the mOR37A gene (right purple arrow). The segments indicated by the blue arrows were only conserved in mouse, dog and opossum; determining their sequence revealed this was part of the putative promoters from mOR37B and olfr71, respectively. Taken together, including opossum in the *Pipmaker* analyses led to the identification of the conserved coding regions or putative promoters, but not any candidate LCRs. Based on this result, it could however not be excluded that relevant LCR segments may be located outside the analysed genomic locus. Therefore the size of the analysed sequence was extended by 200Kb on both sides and these larger elements were subjected to the *Pipmaker* analyses. As shown in Figure 43A and B, besides the conserved segments identified in the previous approach (Figure 35A), one novel conserved segment of about 270 bases emerged, located 137Kb upstream of the mouse locus. The conserved sequence is depicted in Figure 43C; the sequence was downloaded and blasted back against the mouse genome. This analysis indicated that the position of the 270 bases element is in an intron of a non-OR gene, the TlnI-gene. Blasting back the sequence to the genome of each species revealed that whereas in human, mouse and dog the 270 bp segment is one piece, in opossum the 270 bp segment is composed of three comprising 130 bp, 80 bp, 60 bp, respectively. The 130 bp part is located upstream of the cluster, similar to whole segment in the other species. The 80 bp element was located at the beginning of cluster I and the 60 bp piece was located downstream of the cluster. Extensive analyses indicated that the conserved 270 bp segment was not part of any gene exon or part of the identified promoter region of the OR37 genes, suggesting that this 270 bp element may have some alternative role for the OR37 cluster I.

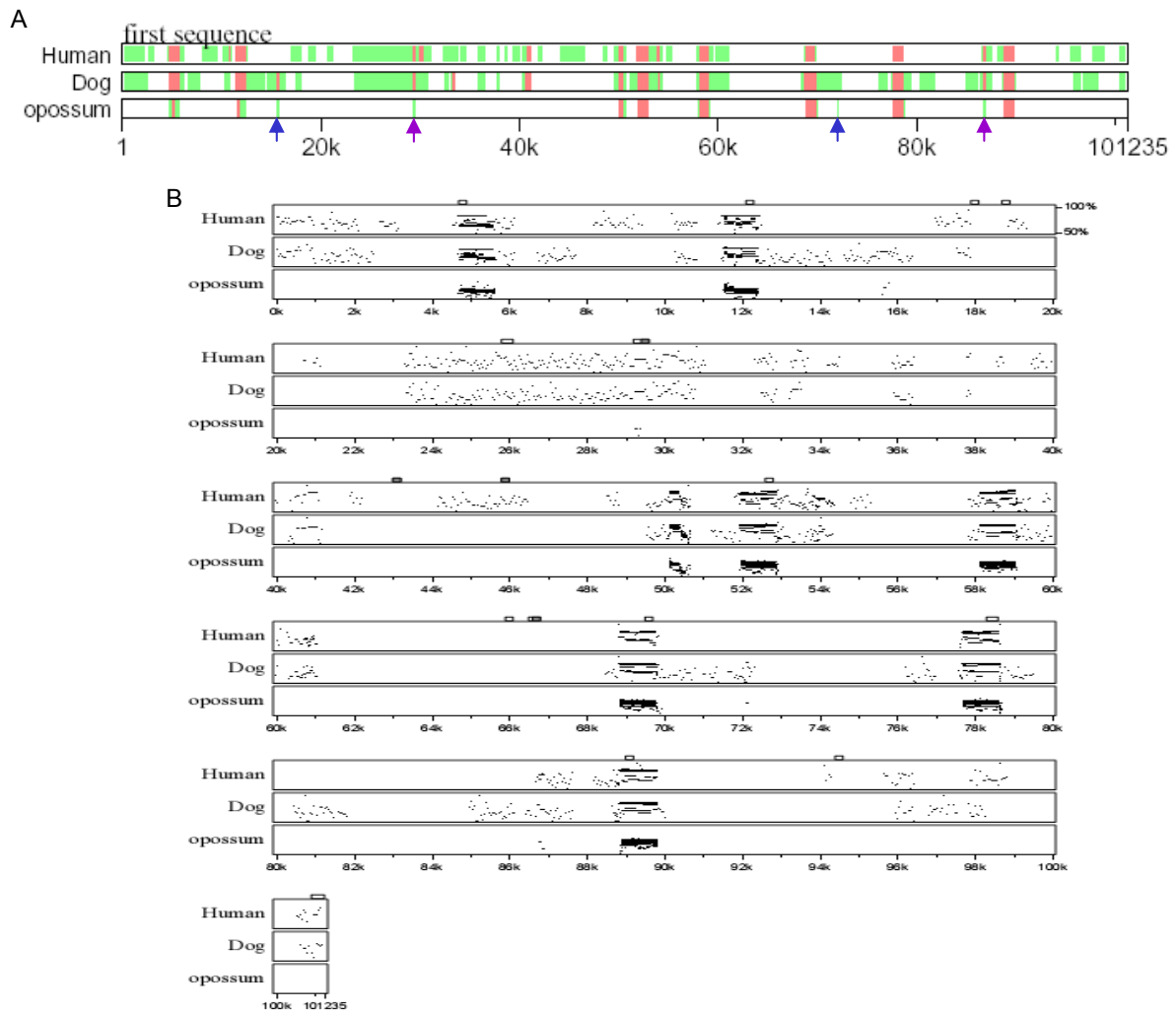


Figure 42 Sequence comparison of OR37 cluster I across mouse, human, dog and Opossum.

A. With mouse cluster I as the reference multi sequence alignments were performed by Pipmaker. The blue arrows indicate the segments only conserved between mouse, dog and Opossum and the purple arrows indicated the sequences conserved in all the species.

B. The detailed segments with similarity from A. All the annotations for the conserved segments were the same as Figure 35B except for the new four segments shown in A by arrows.

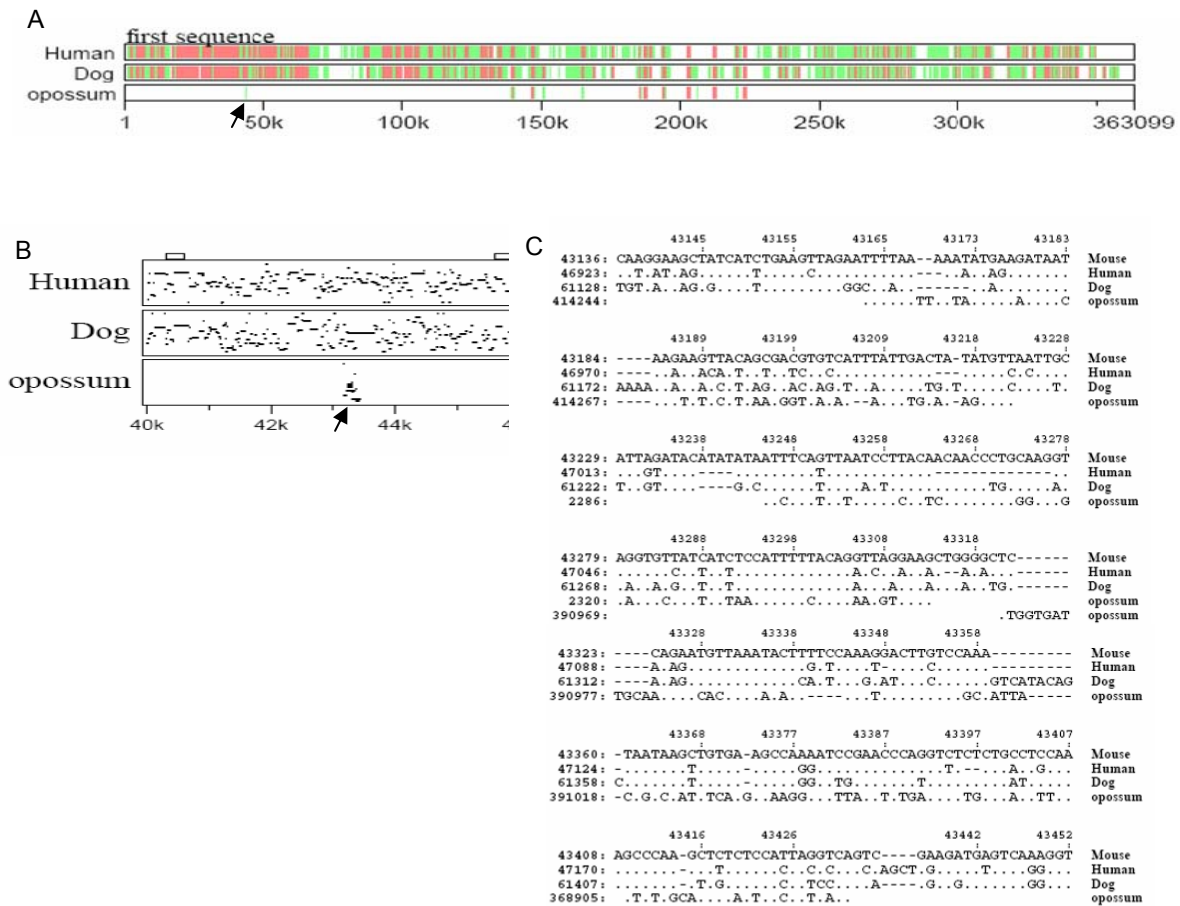


Figure 43 Identification of novel conservations across species by comparison of the expanded cluster locus sequence. In A and B the black arrow indicates the newly identified similarity. C presents the sequence alignment for this segment in detail.

3.4.5 *Pipmaker* analyses of the cluster II locus

Subsequently, a similar multispecies comparison approach was employed for OR37 cluster II. As a first step, the conservation of OR37 cluster II was examined, including the opossum and platypus genomes. The previous study has already demonstrated a syntenic conservation of this cluster among mouse, dog and human; one locus was also found in opossum but nothing is known about platypus. To determine the position of cluster II in the opossum genome and to assess the platypus genome for homologous sequences, cluster II genes (*olfr270*, *olfr271*, *olfr272*, *olfr273*, *olfr274* and *olfr275*) were employed to blast the opossum and platypus genomes. It was found that in opossum, cluster II was positioned on chromosome 6; at the same chromosome as for cluster I (Figure 44). Furthermore, the opossum cluster II was linked to the same downstream genes as cluster II in human, dog and mouse; however, the upstream genes were different (Figure 44). In platypus one OR gene was

found which could be annotated as a homolog of mouse olfr270. Interestingly, this gene was linked to the same downstream genes as cluster II in opossum, dog, mouse and human. The details of mOR37 cluster II conservation are summarized in Table 8.

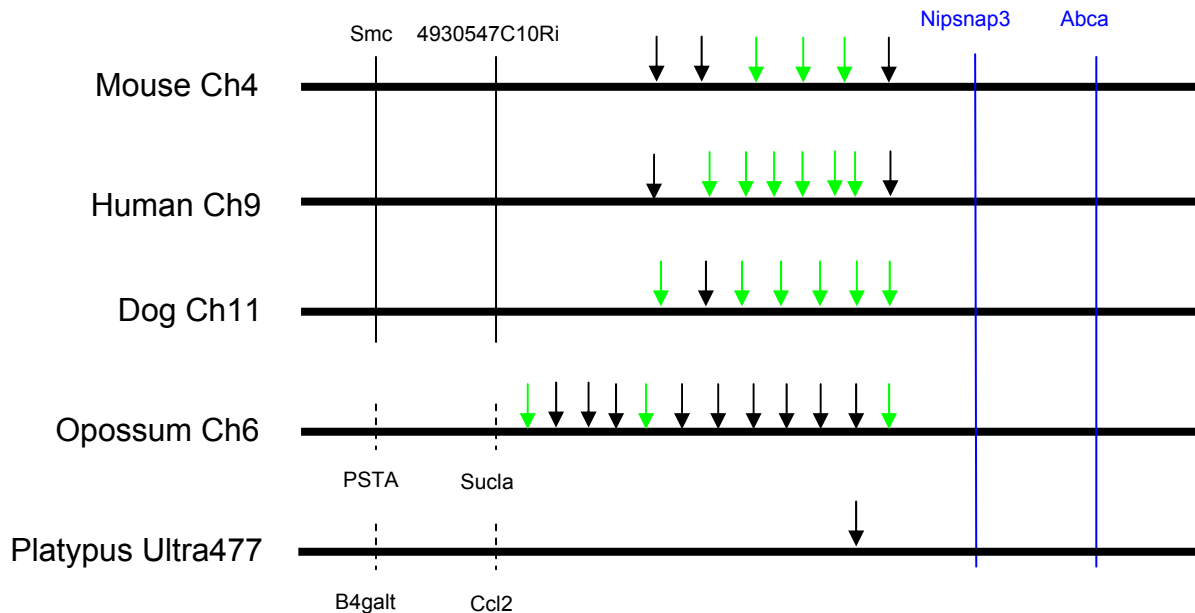


Figure 44 The conservation of OR37 gene cluster II loci in mouse, human, dog, opossum and platypus. The olfactory receptor genes are indicated by arrows, of which the green are the orthologs of mOR37 genes and the black are the OR37 related olfactory receptor genes belonging to subfamily mOR262. The syntenic conservation of the locus is defined by the flanking genes. The black and blue vertical solid lines indicate the upstream and the downstream genes are conserved across species. The dashed vertical lines indicate the variable flanking upstream genes in the opossum and Platypus genome. Ch: chromosome.

Species	Genomic location	OR gene number
Mouse	Ch4: 52633263--53010410	6
Human	Ch9: 105943519--106549790	8
Dog	Ch11: 63376121--63743921	7
Opossum	Ch6: 141275491--141996688	12
Platypus	Ultra477: 455081--550219	1

Table 7 The position and the olfactory receptor gene number of mOR37 cluster II locus across species.

Hoppe et al	Zhang and Firestein	NCBI
---	MOR262-9	Olf270
mOR37F	MOR262-3	Olf271-ps1
mOR37G	MOR262-7	Olf272
mOR37H	MOR262-8	Olf273
---	---	Olf274-ps1
---	MOR262-2	Olf275

Table 8 The composition of mOR37 cluster II and the distinct nomenclature of the corresponding OR genes according to Hoppe et al., 2003; Zhang and Firestein, 2002 and NCBI. The non-OR37 genes are indicated by the dashed lines and the genes are outlined in the order from telomere to centromere.

The conservation of cluster II across the species provided the possibility to search for conserved sequence segments relating to this cluster using the multi sequence comparison approach. The fact, that the upstream genes are variable in opossum and platypus, seemed advantageous for the assignment of any conserved upstream segments. To search for conserved elements the sequences of the cluster II locus from each species extended by 200 Kb on both sides, was downloaded and the repetitive elements were deleted. Subsequently, the sequence comparison was performed with *PipMaker* using the mouse sequence as reference. The results are documented in Figure 45; they confirmed that the OR genes are conserved in all species; furthermore, two conserved elements were found upstream of the cluster. The first conserved element was only found in the 4 species which have a fully

developed cluster II; this element was not found in platypus, which only has a related gene in the corresponding locus (Figure 45B and C). Comparing the sequence of this element confirmed a high degree of conservation among the 4 species; however, the sequence did not show any homology to the 270 bp segment identified for cluster I (chapter 3.4.4). Also the conserved element which was missing in opossum was not related to the cluster I segment.

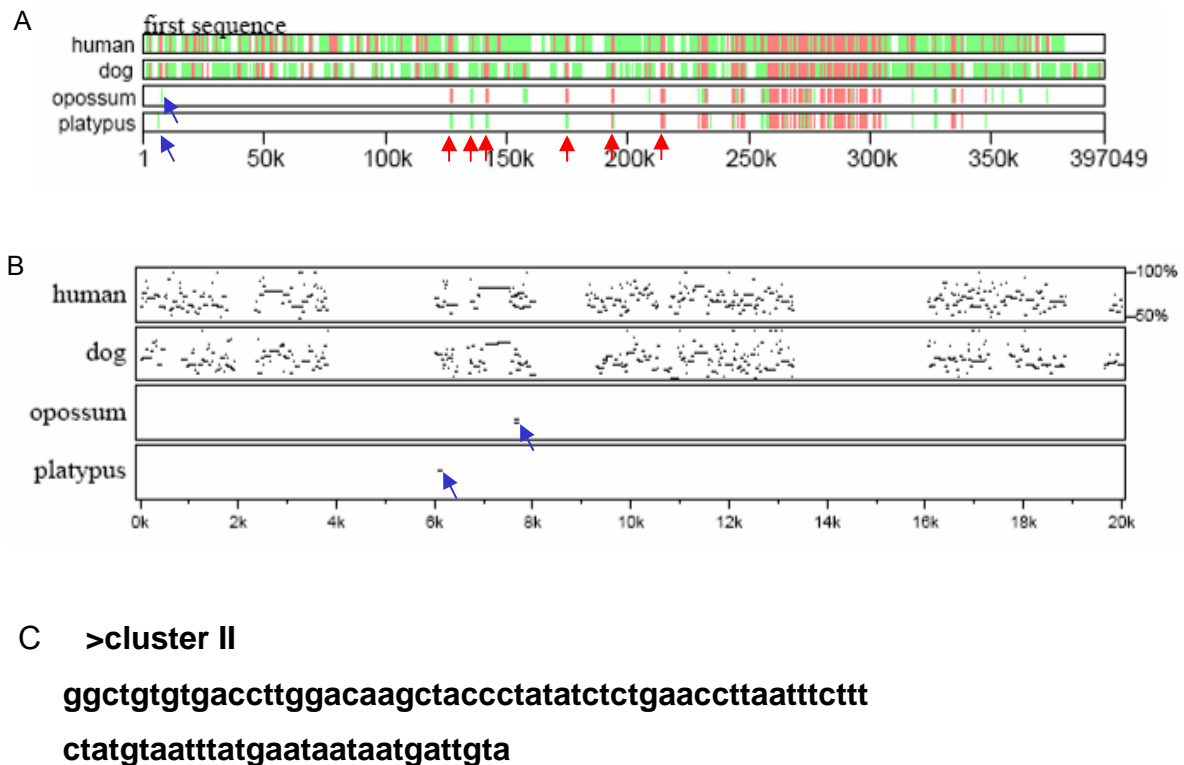


Figure 45 The alignment of OR37 cluster II across species. In A the red arrows indicate the conserved OR genes and the blue arrows in A and B indicate the two conserved elements upstream the cluster. C shows the sequence of the second conserved upstream segment in mouse.

3.4.6 Comparison of the Cluster I and Cluster II locus

The relationship between OR37 cluster I and cluster II suggests that a comparison of the two clusters may be a possibility to identify sequence elements conserved among the two clusters. Furthermore, based on the finding that only one OR37-related gene orthologous to a mouse cluster II gene was found in the platypus genome at a position syntenic to mOR37 Cluster II, it has been speculated that both clusters may have evolved from this single gene in platypus. Gene multiplication and segregation of the gene into two clusters may have occurred before the separation of platypus and opossum, since two clusters exist in opossum genome (Hoppe et al., 2006b).

Therefore, the sequences of cluster I and cluster II from opossum were compared to evaluate if there are conserved sequences besides the coding regions of the OR genes. Surprisingly an almost identical fragment of about 10 Kb was found in cluster I and cluster II; the latter comprising 3 copies (Figure 46). Blasting the opossum genome with this sequence segment did not identify any other hit except these three loci, strongly suggesting the notion concerning a close relationship between cluster I and cluster II loci in the opossum genome. An alignment of cluster I and cluster II across the species did not reveal any conservation of this segment; this was confirmed by blasting the other genomes with this segment. The results imply that this segment appears to be unique for the two clusters in the opossum genome. Interestingly, the different number of this segment in the two clusters was reflected in the number of OR genes, which in cluster II was also 3 fold of that in cluster I. Furthermore, several smaller segments were found to be highly conserved between opossum cluster I and cluster II; no such conservation was observed in cluster I and cluster II of mouse. These observations support the notion that OR37 cluster I may in fact have evolved from cluster II. This view on the other hand would favour the hypothesis that the two clusters may share similar regulatory elements.

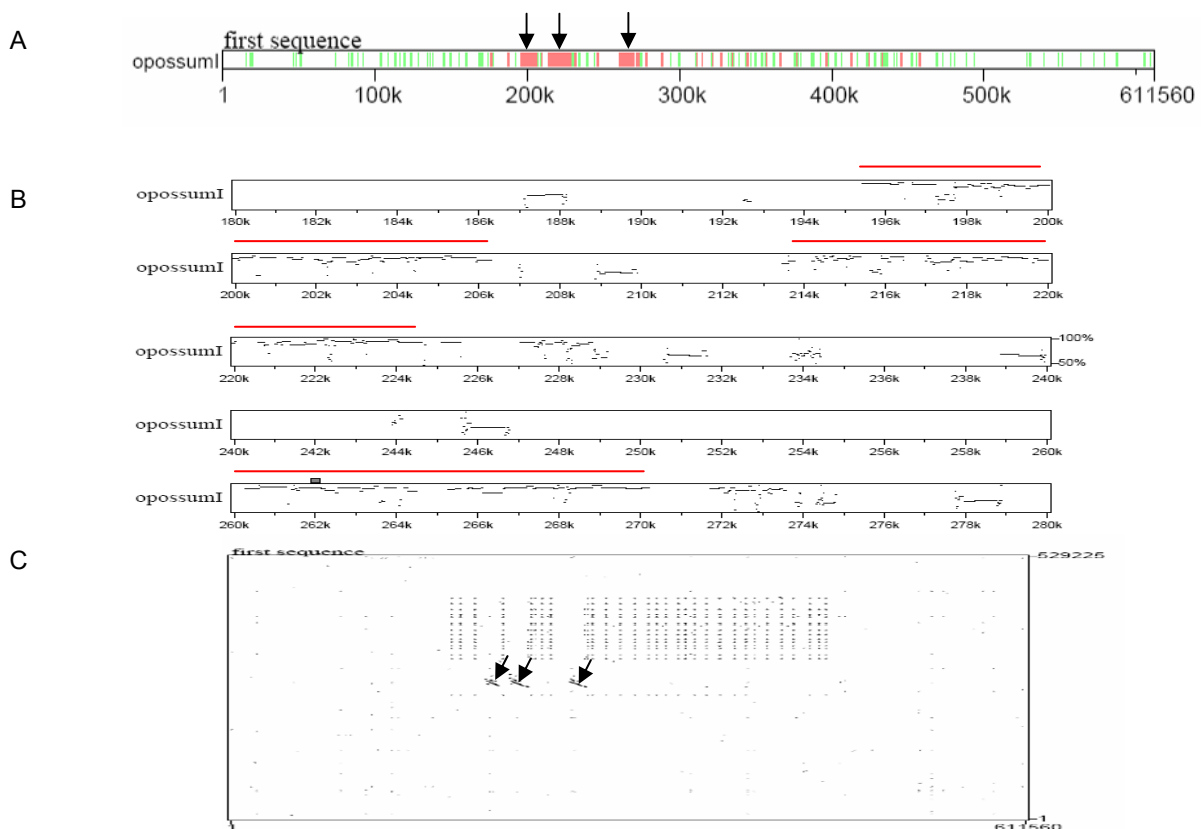


Figure 46 Comparison of the sequences from opossum OR37 cluster I and cluster II.

A. Overview of the alignment of opossum OR37 cluster I to cluster II. The arrows indicate the segments almost identical between cluster I and cluster II, whose location is shown in B by the red lines above the regions.

C. Dot matrix of the alignment as in A.

Consequently, the conserved 270 bp segment identified for cluster I was employed to search for related segments in the cluster II locus. The results for human, mouse and dog are documented in Figure 47A and B. By a comparison of cluster I and cluster II from mouse and human, a conserved segment was not identified; however, by a comparison of the two clusters from dog a homologous segment emerged, which was located in a sector of the cluster II locus. Interestingly, as shown in Figure 47C, the homologous segment in the cluster II locus was located downstream of the cluster. Subsequently, the conserved segment from dog cluster II was downloaded and used to blast to the dog genome; two hits were recognised, with 100% and 99% of identity. Furthermore, the investigation revealed that the two copies of the conserved segment are positioned 8Kb and 60Kb downstream of cluster II, respectively, neither of them located within an exon of a gene (Figure 48).

As a next step, the question had to be addressed whether this segment related to cluster II may be unique for dog or may be present in the cluster II locus of all species. To approach this question, an alignment of the cluster II locus from all species was performed using dog cluster II as reference sequence. The results shown in Figure 46A and B demonstrate that the segment was indeed found in all the species. By a more detailed analyses of the conserved segment found in all species (Figure 49C) a sequence motif of 111 base pairs emerged. This 111 base pair segment was then downloaded and a comparison with the alignment in Figure 46 revealed that its homologous counterpart in dog cluster I comprises 125 bases (Figure 50A and B). Comparing this segment of 125 bases in the alignment of cluster I from all species (Figure 43) revealed that the corresponding part in mouse cluster I comprises 115 bases (Figure 50C). In a summary, a segment related to the 270 base pairs element of cluster I was found for cluster II; the elements for the two clusters shared a conserved sequence motif of 115 base pairs.

A >202 bases from cluster I conserved between dog cluster I and II
 tacgtcatttcatttaaaccttacaacaacctggcaagataagtatgattattccattttacagattaag
 aaactgaggtgcaagatgtaaatactcattcaaggatcttctc caaagtcatacagctaataagtt
 gtgaagccaggattggaacctatgtctcttgattccaagcccaagttgtctccactatcc

>counterpart of 202 bases from dog cluster II
 tacctcatttcattatcgttacaaccctaactaggagtaacattatcaccaatcaacaacaagggaa
 ttgagctcagagagactaaggaacttgccaaggtcacacagccaataagtagcagcaccagaa
 ttccaaccaggctgtgtgtctccaaactcagactctctgcattaccactt

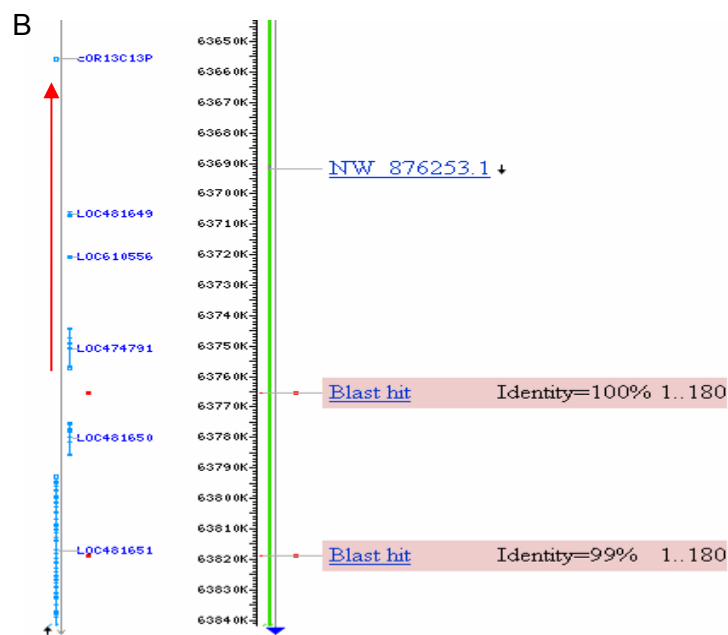


Figure 48 The annotation of the conserved segments in dog cluster I and cluster II.

A. Sequences of the segments in dog cluster I and cluster II.

B. Blasting the dog genome with the second sequence identified two hits downstream of dog mOR37 cluster II. The red arrow shows the OR37 Cluster II in dog genome.

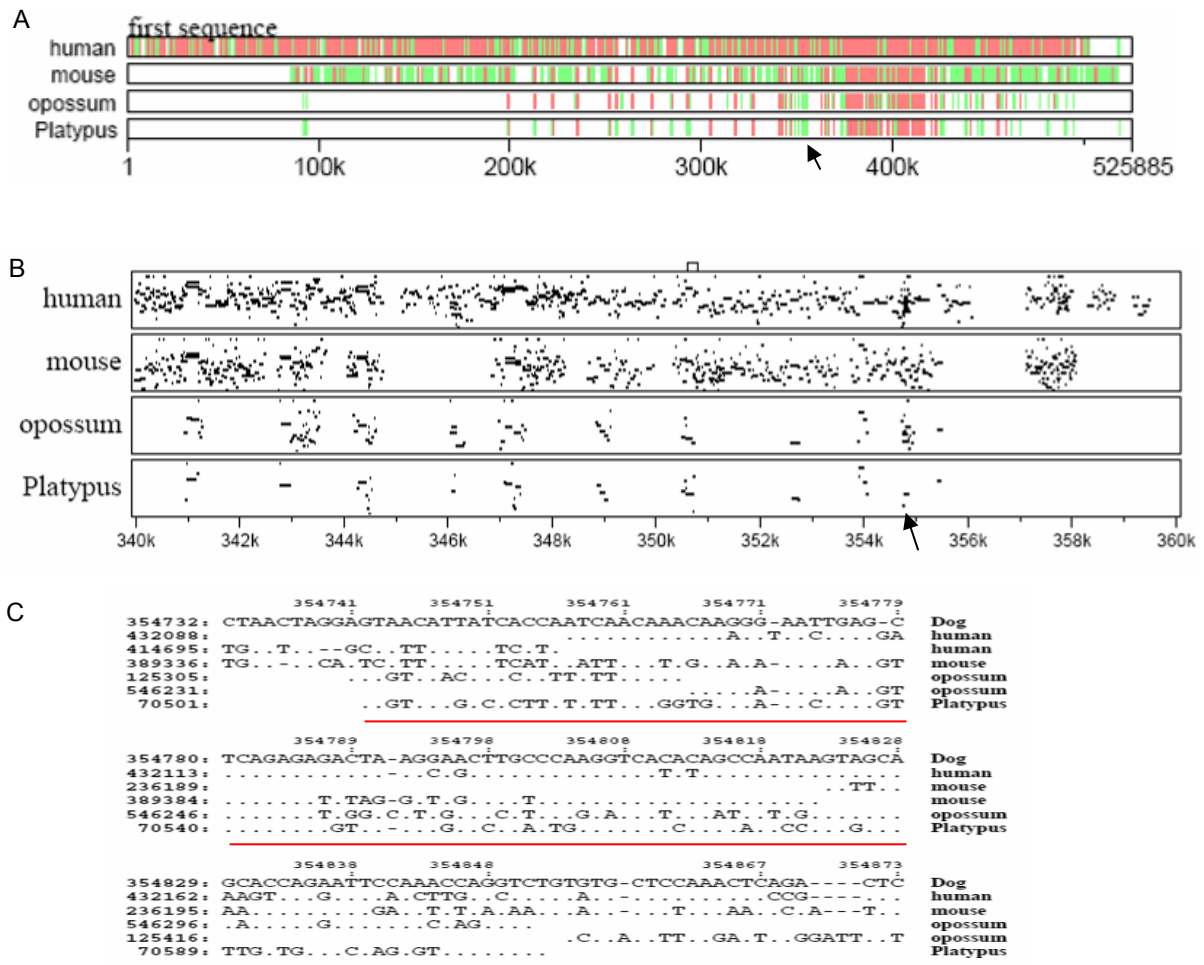


Figure 49 The alignment of cluster II using dog as reference.

A and B show the overview of the alignment. The arrow indicates the alignment of the dog cluster II segment from Figure 44 with other species.

C. Details of the region indicated by the arrow. The red line below marks the sequence conserved in all species.

- A >111 bases from dog cluster II
 gtaacattatcaccaatcaacaacaaggggaattgagctcagagagactaagg
 Aactgccaaggtcacacagccaataagtagcagcaccagaattccaacca
 ggtct
- B >125 bases from dog cluster I
 gtatgattatttcattttacagattaagaaactgaggtgcaaagatgttaaatactc
 attcaaggatcttctccaaagtcatacagctaataagttgtgaagccaggattgga
 acctatgtct
- C >115 bases from mouse cluster I
 gtgttatcatctccattttacaggtaggaagctggggctccagaatgttaaatact
 ttccaaaggactgtccaaataataagctgtgaagccaaaatccgaaccagg
 tct
- D >270 bases from mouse cluster I
 ttttaaaaatatgaagataataagaagttacagcgacgtgtcatttattgactatag
 ttaattgcattagatacatatataatttcagttaatccttacaacaaccctgcaaggta
 ggtgttatcatctccattttacaggtaggaagctggggctccagaatgttaaatact
 ttccaaaggactgtccaaataataagctgtgaagccaaaatccgaaccagggt
 ctctctgcctccaaagcccaagctctctccattagggtca

Figure 50 The sequence conserved in both mOR37 cluster I and II.

- A. The sequence indicated by the red line in Figure 49C and B show its homologous counterpart in dog cluster I.
- C. Corresponding part of the segment of 270 bases according to B.
- D. The conserved segment of 270 bases upstream of mouse mOR37 cluster I.

3.4.7 Comparison of the conserved OR37 segment and the H element

As a next step, attempts were made to explore if the identified segments of the OR37 cluster may share sequence motifs with the H element for OR28. Locus control regions are characterised by DNase I hypersensitive sites (LCR-HS). Previous studies have shown that Zinc finger proteins can facilitate the formation of a hypersensitive site and subsequently bind to a specific site comprising a characteristic sequence motif (AT)GATA(AG), which was considered as predominant motif within the LCR (Davies et al., 2004; Vakoc et al., 2005; Im et al., 2005; Layon et al., 2007). To

examine whether such motifs may also be present in the conserved segments found for the OR28 cluster and OR37 cluster, a screening approach was performed using program "Match and Patch" against the transcription factor binding sites database TRANSFAC 7.0. In Figure 51 the motifs obtained for each segment are depicted. In both segments two kinds of motifs, GATA and OCT-1, were identified in the H element as well as in the OR37 cluster I flanking segment; additional motifs were found in each sequence. When the 115 sequence conserved in both, OR37 cluster I and cluster II loci was analysed, it was found, this element only contained the GATA motif. This finding indicates that the conserved OR37 element does contain the GATA motif, which seems to be the typical motif in the LCR-HS formation. Taken together, both conserved sequence elements, the H element for the OR28 cluster as well as the segment of the OR37 clusters does comprise a motif which has been attributed to contribution to the formation of DNase I hypersensitive chromatin in locus control region.

A	matrix identifier	position (strand)	core match	matrix match	sequence (always the (+)-strand is shown)	factor name
	V\$TGIF_01	13 (-)	1.000	0.957	tttTG&CAagt	TGIF
	V\$GATA1_05	43 (+)	1.000	0.973	tctGATAAga	GATA-1
	V\$GATA2_02	43 (+)	1.000	0.977	tctGATAAga	GATA-2
	V\$GATA_C	45 (+)	1.000	0.972	tGATAAgatga	GATA-X
	V\$CP2_01	54 (+)	1.000	0.901	gaacaaCCC&G	CP2
	V\$CP2_01	103 (-)	0.987	0.929	CTGGTtcatgc	CP2
	V\$OCT1_Q6	106 (+)	0.883	0.806	gttcatGCTA&catc	Oct-1
	V\$CDPCR1_01	169 (+)	0.794	0.836	cATTA&ttgc	CDP_CR1
	V\$HNF4_01	173 (-)	0.752	0.807	aattgccCTCTG&tctcaa	HNF-4

B	matrix identifier	position (strand)	core match	matrix match	sequence (always the (+)-strand is shown)	factor name
	V\$EVI1_04	5 (+)	1.000	0.748	aaaaat&tgAAGATa	Evi-1
	V\$GATA1_03	11 (+)	1.000	0.981	atgaaGATA&taag	GATA-1
	V\$GATA1_06	13 (+)	1.000	0.972	gaaGATA&Aata	GATA-1
	V\$GATA2_02	13 (+)	1.000	0.970	gaaGATA&Aata	GATA-2
	V\$GATA2_03	13 (+)	1.000	0.970	gaaGATA&Aata	GATA-2
	V\$GATA3_03	13 (+)	0.958	0.937	gaaGATA&Aata	GATA-3
	V\$HNF1_C	37 (-)	0.874	0.847	cgtgtc&atttaTTG&Ct	HNF-1
	V\$OCT1_Q6	52 (+)	0.792	0.825	ctatatGTTA&ttgc	Oct-1
	V\$OCT1_Q7	53 (+)	1.000	0.997	tatatgtTA&ATT	Oct-1
	V\$S8_01	56 (-)	1.000	0.991	atgtTA&ATTgcattag	S8
	V\$CHOP_01	58 (-)	1.000	0.920	gttaaTTGC&Atta	CHOP-C/EBPalpha
	V\$NKX25_02	58 (+)	1.000	0.931	gtTA&ATTg	Nkx2-5
	V\$OCT1_Q6	58 (-)	0.764	0.834	gttaaTTGC&attaga	Oct-1
	V\$BRN2_01	73 (+)	1.000	0.937	tacatat&ATA&Tttca	Brn-2
	V\$VMYB_01	85 (-)	0.938	0.919	ttC>Ta&at	v-Myb
	V\$HNF1_01	88 (+)	1.000	0.919	agTTA&Tc&ctta&caa	HNF-1
	V\$GATA1_05	120 (-)	1.000	0.984	tgTTAT&C&atc	GATA-1

C	matrix identifier	position (strand)	core match	matrix match	sequence (always the (+)-strand is shown)	factor name
	V\$GATA1_05	2 (-)	1.000	0.984	tgTTAT&C&atc	GATA-1

Figure 51 The motifs identified in the conserved segment.

- The 202 bases of the mouse H element.
- The 270 bases of segment from mouse cluster I.
- The 115 bases conserved between cluster I and cluster II from the 270 bases of segment. The red arrows indicate the motif existing in all three sequences and the green arrows indicate the common motif in A and B.

4. Discussion

The interaction of transcription factors with the promoters of genes plays a crucial role in initiating and regulating gene expression. Members from two families of transcription factors, the O/E proteins (Olf/EBF) and the LIM-homeodomain proteins, have been demonstrated to be involved in governing the differentiation of olfactory sensory neurons (Wang et al., 2004; Hirota et al., 2004). At least four O/E proteins (O/E-1, O/E-2, O/E-3 and O/E-4) and one LIM-homeodomain protein (Lhx-2) are expressed in the olfactory epithelium (Wang et al., 1993; Wang et al., 1997; Wang et al., 2002; Wang et al., 2004; Hirota et al., 2004). O/E proteins are characterised by a conserved olf-1 binding domain at the N-terminal region, an IPT domain which binds cofactors, a helix-loop-helix domain for dimerization and a transactivation domain at the C-terminus (Wang et al., 1993; Hagman et al., 1993; Liberg et al., 2002). LIM-homeodomain proteins are featured by the so-called LIM domain which binds to the cofactor and a homeodomain which binds the TAAT sites within the regulatory elements of the genes (Hobert et al., 2000). Our lab has previously demonstrated that a few transcription factors including O/Es and Lhx-2, physically bind to the promoter region of OR37 genes (Hoppe et al., 2003b; Hoppe et al., 2006b); this was also reported for some other OR gene promoters (Rothman et al., 2005; Hirota et al., 2004; Vassali et al., 2002). However, in none of the cases it was evaluated if such binding may lead to an expression of the OR genes. In the present study, it was demonstrated that an interaction of O/E or/and Lhx-2 with the mOR37 promoter indeed elicited the expression of a reporter gene in HEK cells (Figures 2, 10 and 13), indicating a functional interaction between transcription factors and promoter. The data suggest a functional effect of O/Es on OR gene promoters in general; since they not only induced the reporter gene expression via the promoter of the OR37 gene but also via promoters of other class II as well as class OR I genes. Such kind of activation monitored by means of *in vitro* systems was also reported for promoters of genes encoding the olfactory marker protein (OMP), the type III adenylyl cyclase (ACIII), the olfactory cyclic nucleotide-gated channel (OCNC) and the olfactory G-protein (Golf) (Wang et al., 1997; Wang et al., 1993; Kudrycki et al., 1993; Kudrycki et al., 1998; Liberg et al., 2002). For the precursor cells and for the still immature olfactory sensory neurons in the epithelium which do not express OR genes it was found that the protein Roaz may play an important role. Roaz is a zinc finger protein which binds to O/E-1 and apparently prevents the physical and functional interaction

between O/E-1 and promoters; thus, assuring the silencing of genes in immature cells which are only expressed in mature OSN (Tsai and Reed, 1997). For the transcription factor Lhx-2 a quite distinct scenario emerged; an activating effect was only found for some OR promoters but not for others (Figure 2), suggesting a more selective role. This observation is in line with the recent finding that Lhx-2 has a differential impact on the expression of class I and class II receptor genes (Hirota et al., 2007). It was demonstrated that a deficiency of Lhx-2 precludes expression of class II ORs but not of class I ORs.

These results seem to be contradicting the finding that Lhx-2 does activate OR11-1 and OR18-2 genes in the heterologous system (Figure 2); however, the additive effect of O/Es and Lhx-2 on luciferase expression governed by OR promoters (Figure 13) indicate that the level of OR expression may be regulated by an interplay of both factors; the contribution of each may vary between class I and class II OR-genes. This view is supported by the observation that e.g. mOR37C was very strongly activated by the O/Es whereas Lhx-2 had only a small effect; in contrast, for mOR18-2 the activation by O/Es was moderate but the effect of Lhx-2 was quite strong (Figures 2, 7 and 9). This finding is quite surprising, as all class I and class II OR genes do have Lhx-2 binding sites in their promoter region.

It is of particular interest to elucidate the effect of O/Es and Lhx-2 on the expression of OR37 genes due to their unique expression pattern in the olfactory epithelium (Strotmann et al., 1992). The combination of motif I and IV contains a homeodomain binding site and an olf-1 binding site is characteristic for all promoters of mOR37 genes, suggesting a significant role with respect to gene selection and gene expression pattern (Hoppe et al., 2006b).

In view of the mOR37 promoters, they actively primed the reporter gene expression when the critical transcription factors were present in the heterologous system. But this did not reflect its function *in vivo* especially in the terms of to which extent and how the promoter drives the mOR37 gene expression and maintains the expression and projection pattern of mOR37 expressing neurons. To address this question a transgenic strategy was employed.

In the present study it has been demonstrated that a short transgenic construct spanning less than 10 kb of genomic DNA surrounding the coding region of the OR gene mOR37C (*mOR262-12*; Olfr157) was sufficient to specifically drive its expression in distinct populations of OSNs segregated in the central patch of the

turbinates. Thus, the unique spatial segregation of the OSNs expressing mOR37 genes was largely reproduced. Moreover, the cells expressed the transgene in a mutually exclusive and monoallelic fashion as previously shown for the endogenous mOR37 genes (Strotmann et al., 2000). These results are also in line with the findings obtained when transgenes of OR genes which are expressed in a zonal pattern were studied (Qasba and Reed, 1998; Vassalli et al., 2002; Lewcock and Reed, 2004; Rothman et al., 2005). All these results support the concept that OR genes are very compact units. Their compactness has been considered as an important feature for facilitating the extraordinary expansion of the mammalian OR gene repertoire (Zhang and Firestein, 2002). The mOR37C transgene construct employed in this study comprised only 358 bp upstream of the transcription start; this stretch of DNA contained the sequence motifs which are highly conserved in the upstream region of those OR genes which are expressed in the clustered pattern (Hoppe et al., 2000; Hoppe et al., 2006). Our previous studies using bioinformatic tools revealed that these motifs form a unique array in the mOR37 gene promoters; they were found to be associated with binding sites for homeodomain transcription factors and O/E binding sites which were also found in large variety of other OR gene promoters (Vassalli et al., 2002; Rothman et al., 2005; Michaloski et al., 2006). From these findings it was concluded that a particular combination of such DNA motifs is involved in regulating the spatial organization of OR gene expression. The result that in almost all generated mouse lines the transgene was expressed only in such OSNs which were segregated in the characteristic clustered pattern exclusively in the central patch of the turbinates strongly suggests, that the upstream motifs contain the relevant information to mediate this unique expression pattern. This result confirms and extends our previous notion that the promoter region of OR genes expressed in a defined topographic manner contains distinct sequence motifs (Hoppe et al., 2006b). In fact, studying a transgene of the OR gene M71 which is expressed in the dorsal zone of the epithelium, revealed that upstream elements within a DNA-stretch of a few hundred bps are capable to recapitulate the zonal expression pattern of M71 (Rothman et al., 2005). Furthermore, mutations within the promotor region of M71 strongly affected the spatial localization of OSNs expressing this OR gene. Altogether, the current data indicate that short proximal promoter elements of OR genes seem to govern their expression, although, at this point it cannot completely be ruled out that other DNA motifs in the transgenic construct may also contribute to

the control of expression. However, in this context it is interesting to note, that so far no evidence was found that the intron sequences may have any essential information for this aspect (Vassalli et al., 2002).

The results obtained for the mOR37C transgene indicate that the control elements can act autonomously, i.e. outside the original gene cluster, which implies that no locus control element in *cis* is required. Thus, the spatial expression pattern of OR genes seems to be regulated proximally. This concept is consistent with very recent data which indicate that the choice of a single OR gene by an OSN is mediated by a mechanism which requires an association of the OR gene promoter with a DNA element located in *trans*, which seems to exist as a singly copy in the genome (Lomvardas et al., 2006); such a principle would indeed make the choice of an OR gene largely independent from its location in the genome.

In spite of their relative autonomy, under certain conditions even the compact OR genes seem to be quite sensitive to the genomic region where a transgene is integrated. Most obvious is the different number of transgene expressing cells in the various mouse lines (Figure 16); such a phenomenon has also been described for MOR23 and M71 (Vassalli et al., 2002; Rothman et al., 2005). The observation that also the staining intensity of transgene expressing cells was different in the various lines indicates that the transgene integration in different genomic loci affects the level of transgene expression in a given cell. For line #7, in addition to the cluster, ectopically positioned cells expressing the transgene were found; interestingly, however, even in this case the cells were not randomly distributed throughout the epithelium but segregated in the medial zone, typical for other OR-types, like e.g. P2. The basis for this phenomenon is unclear, however, it seems conceivable that the transgene was integrated in the vicinity of an OR gene cluster which exerts a significant influence on the transgene. Similar phenomena have been observed in previous studies. It was found that a transgene of OMP which was integrated within an OR gene cluster was not expressed in all mature OSNs but rather in a spatial pattern typical for the OR genes of that cluster (Pyrski et al., 2001). Furthermore, mutations in the promoter region of the endogenous M71 gene had less severe effects on the expression pattern than the same mutation in the promoter region of a transgene (Rothman et al., 2005). Both observations have led to the conclusion that within certain gene clusters control elements exist which are capable to affect promoters in their proximity.

In all mouse lines enough cells expressed the mOR37C transgene to trace their axonal projection into the olfactory bulb (OB). It was found that in all cases the axons converged onto defined glomeruli. Based on the concept that only OSNs expressing the same OR gene send their axons onto common glomeruli (Mombaerts et al., 1996; Wang et al., 1998), this result strongly suggests that the respective OSNs express only mOR37C. Moreover, it was found that only one of the two alleles from the transgene was chosen by an individual cell (Figure 24), indicating that the transgene contained all necessary information to ensure a monoallelic expression of a receptor and suggests that the feedback mechanisms which insure that expression of only one OR gene is maintained in a given cell (Serizawa et al., 2003; Lewcock and Reed, 2004; Shykind et al., 2004) can act on the transgene.

In most of the lines, all OSNs expressing the mOR37C transgene were clustered within a patch of the epithelium, like their endogenous counterparts and the axons of both populations co-converged, as expected for OSNs expressing the same OR gene. However, in some of the lines the degree of co-convergence was not as perfect as seen for the two OSNs populations expressing one of the two wild type alleles (Strotmann et al., 2000). This is consistent with the current model that not only the type of OR, but also the level of OR protein may be a critical determinant for guiding the axon to the specific target in the OB (Feinstein et al., 2004; Imai et al., 2006). It is quite possible that the level of expression from the transgenic locus may not precisely match the wild type locus and therefore the axons might be slightly mistargeted.

A more pronounced difference of projection from OSNs expressing the mOR37C transgene was seen in line #7, in which about 4 – 6 glomeruli were targeted by these axons (Figure 27 and 28). This projection pattern correlated with the dispersed distribution of the corresponding OSNs and implies that those OSNs which were ectopically positioned in the olfactory epithelium form novel glomeruli. Recent studies have demonstrated that in addition to the OR expressed by a given OSN, also the position of the respective OSNs in the epithelium is a critical determinant of the projection pattern (Levai et al., 2003; Feinstein et al., 2004; Miyamichi et al., 2005). It has been discussed that this may be due to a differential distribution of guidance molecules that act in concert with the OR in the targeting process. The additional glomeruli formed by the axons of transgene expressing cells were located on the ventral side of the bulb, rather close to the original site. Interestingly, they sometimes

contained also fibers from OSNs expressing the endogenous gene, an observation which has previously also been made for ectopically positioned MOR23 glomeruli (Vassalli et al., 2002). The phenomenon that axons from OSNs expressing the endogenous gene are re-routed to those novel positions has been viewed in favor of the idea that the receptor protein itself, which is present in the axonal processes (Feinstein et al., 2004; Barnea et al., 2004; Strotmann et al., 2004), may via homophilic interactions mediate the coalescence of these axonal processes into the same glomerulus (Feinstein and Mombaerts, 2004).

The results demonstrate that the unique expression pattern of the mOR37C gene exclusively in OSNs segregated in a small area of the central turbinate is recapitulated by the short transgene. This finding confirms the concept that most of the control elements governing the restricted topographic expression of this gene are located in a rather short stretch of DNA just upstream of the transcription start site, however, the delayed initiation of the transgene expression, copy number independence of transgene expression and ectopic position of OSNs expressing the transgene indicate the involvement of other regulatory elements. The most popular transgene which is able to completely fulfill the features of the endogenous counterpart is the beta-globin transgene containing both the proximal and distal regulatory elements.

The comparative bioinformatic approach employed in this study trying to identify potential regulatory elements for the OR37 genes which may be located far distant from the gene cluster, followed the strategy that has been successfully used in search for the Locus Control Region (LCR) of the beta-globin genes (Buhler, 2001; Li et al., 1999; Hardison et al., 1997) as well as for potential LCRs of other clustered genes (Regales et al., 2003). The conserved segment for cluster I identified by this approach (Figure 43) is likely to be an LCR-element for the mOR37 genes of this cluster. This finding demonstrates again the feasibility to find LCR-like elements by an extensive comparison of sequences from distantly related species. As in other cases, further experimental evidence is needed to unequivocally prove that the identified segment is in fact the LCR for cluster I of OR37 genes, employing approaches like for the H-region (Serizawa et al., 2003) or the LCR of beta-globin genes (Grosveld et al., 1987). In the meantime, however, additional observations support the view that the identified genomic segment is in fact a relevant element for the OR37 genes; this is based on the existence of two highly related clusters of

OR37 genes. Sequence comparison between cluster I and cluster II led to the identification of a conserved segment also for the cluster II locus. This result strongly supports the notion that the conserved segment is indeed part of the OR37 cluster and indicates that a similar element may operate as “LCR” for each of the two OR37 clusters. In this context, it is interesting to note that a similar bioinformatic approach, an extensive sequence comparison within and across species has also contributed to identify the conserved promoter of the mOR37 genes (Hoppe et al., 2003a; Hoppe et al., 2006b). The conservation of both the promoter of OR37 genes and the candidate locus control region for the OR37 gene clusters points to a possible interplay of these two elements in regulating the expression of individual OR37 genes. Furthermore, the data confirms and extends the notion that the OR gene clusters are syntenically conserved even in distantly related species, thereby supporting the concept that duplication and translocation of genomic DNA gave rise to novel but highly related gene clusters during evolution (Young et al., 2002; Niimura and Nei., 2005; Aloni et al., 2006). Thus, the comparative approach seems to be an efficient strategy in search for genomic elements which are related to a genomic locus.

Comparing the conserved element of the two OR37 clusters revealed that for cluster II the position of this element is much closer to the actual group of genes than for the element of cluster I; for cluster II the distance was about 10 Kb; for cluster I about 130 Kb (Figures 43, 48 and 49). The distance between the LCR and the respective genes may have immediate functional implications; this has been reported for the LCR controlling of the beta-globin gene expression (Dillon et al., 1997). Moreover, a recent study analysing mOR28 transgenes with the H element positioned at different distances to the mOR28 gene cluster have shown that in cases when the H element was positioned closer to the cluster, the most upstream located OR gene was expressed in a significantly higher number of cells than in controls, whereas the more downstream located genes were expressed in a lower number of cells (Serizawa et al., 2003). If such kind of effect is also operating for the OR37 gene it would be interesting to figure out whether the gene in cluster II which is most closely located to the conserved element is expressed in more cells than the corresponding gene in cluster I. This could give important insight, since the position effect of distantly located regulatory elements on the transgene expression has also been challenged and considered as a phenomenon of the transgene, rather than a reflection of the wild type situation since a recent study has shown that a functional

interaction brings together the regulatory elements which are far apart from each other, even on different chromosomes (Lomvardas et al., 2006).

In human, mouse and dog the potential “LCR” was found to be a continuous segment of about 300 bp; whereas in the OR37 locus of opossum the conserved segment consists of three distinct parts which are located upstream, within and downstream of the cluster, respectively (Figure 43). During evolution three parts apparently fused into one segment giving rise to the conserved elements upstream of the OR37 cluster I in dog mouse and human. In spite of the overall similarity of the segments for cluster I and for cluster II, the three subelements can still be recognised. The sequence element of the LCR for the beta-globin genes also represents one continuous upstream region of the locus in goat, mouse and human (Tuan et al., 1985; Moon and Ley., 1990; Li et al., 1999), whereas in chicken one LCR segment is positioned within the beta globin gene cluster and thus separated from the ones upstream of the cluster (Reitman et al., 1990). Now it will be interesting to explore whether elements of the LCR for the beta-globin genes in species more ancient than chicken may also be separated as the potential LCR for the OR37 cluster I in opossum. However, it seems to be a feature of several other LCRs that elements are spread within the whole cluster locus (Bonifer et al., 1990; Aronow et al., 1992; Baker et al., 1999). Whether the similarity to the LCR for beta-globin genes indicates that the identified conserved segment is the only LCR for OR37 cluster I requires further investigation. Although the two identified potential “LCR” of the OR37 gene clusters show an overall sequence similarity, a comparison of the three subsections revealed that the first section of the “LCRs” for OR37 cluster I and cluster II was especially conserved, suggesting a particular role of this region. This notion is supported by previous studies, indicating that the unique part of a conserved promoter or enhancer region is involved in the specificity of gene expression (Cole, 1983; Cheung et al., 2006). Whether this is also the case for the conserved elements of the OR37 cluster remains to be elucidated.

The conserved segment of mOR37 cluster I is neither part of the promotor nor an exon of any known gene. Two predominant binding sites for transcription factors, GATA and Oct-1, were found within both segments and the one conserved in the mOR28 locus (Figure 51). GATA binding sites exist in DNase-I hypersensitive elements of the LCR for the beta globin gene locus (Jarman et al., 1991; Caterina et al., 1994). It has been shown that upon binding of the zinc finger protein GATA-1,

GATA-2 or GATA-3, the generation of open chromatin was elicited; this is considered as a prerequisite for the high level expression of the beta-globin gene (Joulin and Richard-Foy., 1995; Cao and Moi., 2002; Layon et al., 2006). Interestingly, in a recent study it has been shown that a member of this zinc finger family, GATA-2, is expressed in the olfactory epithelium (Shetty et al., 2005). A second site within this segment is a binding site for the ubiquitous transcription factor Oct-1, which plays an important role in controlling the expression of many genes. In one particular case, it has been shown that Oct-1 and GATA-1 together regulate the expression of a hemoglobin-stabilizing protein by binding to the corresponding motifs in the promoter region (Ryan et al., 2000; Liu et al., 2005); furthermore, there is also evidence for an involvement of both GATA-1 and Oct-1 in governing the LCR-function (Molet et al., 2002). An important binding site in the LCR of the globin gene cluster for NF-E2 is missing in the identified 'LCR' of mOR37 clusters. In this context, it is interesting to note that the transcription factor NF-E2 is only expressed in the erythroid cell lineage (Kotkow and Orkin., 1995). Consistently, there is no evidence for this site in the promoter of the extensively investigated mOR23, M71 and mOR37 genes (Vassali et al., 2002; Rothman et al., 2005). Besides the common GATA and Oct-1 sites, several TF binding sites were identified which seem to be distinct for the 'LCR' of the mOR37 cluster and not present in the H region of them OR28 cluster. The expression of the mOR28 transgene was determined by both the H element and the promoter region (Serizawa et al., 2003). If this is also the case for the mOR37 genes it is conceivable that TF binding sites in the OR37-'LCR' which mediate the interaction with the promoter should be different from the H element; some of the conserved motifs of mOR37 gene are only shared by OR genes exhibiting the same expression pattern on the OE (Hoppe et al., 2006b).

The distantly located regulatory elements are enhancer or locus control region. The notion that the identified conserved segment may act as an enhancer seems to be supported by the observation that the level of expression for transgenes which do not include the potential LCR-region, in our transgenic mouse lines #1 to #6 was significantly reduced (Figure 16), however, the elevated level of expression of the transgene in line #7 cannot readily be reconciled with this concept. A previous study has demonstrated that it is a locus control region rather than an enhancer that is capable of eliminating the negative and positive effect from the transgene locus and enables the transgene expression in a locus independent and copy number

dependent way (Grosveld et al., 1987; Milot et al., 1996). Furthermore some element within the locus control region also exhibits an enhancer activity (Fraser and Grosveld., 1998; Li et al., 2002). Thus, the variable effect of the integration site on the mOR37C transgene expression implies that it is likely the LCR rather than an enhancer that was missing from the transgene. This notion may also be supported by the observation that the expression of transgenes was similarly delayed during embryonic development for all the lines (Figure 20). It has been shown that the elements for controlling the precise expression of distinct globin genes during development are located within the LCR (Li et al., 1998; Huang et al., 2000; Li et al., 2004; Kim and Dean., 2004). Whether the identified segment in fact acts as the locus control region for the mOR37 gene cluster remains to be elucidated by further transgenic and knock-out approaches.

5. Summary

The genes of the OR37 family are clustered in two loci (cluster I and cluster II) on mouse chromosome 4. These genes encode distinct olfactory receptors (ORs) which are characterised by an insertion of six amino acids in the third extracellular loop and moreover, these receptor types are only expressed in cells which are segregated in a small patch on the central nasal turbinate. As first steps to unravel the molecular basis of this unique topographic expression pattern previous studies have led to the identification of highly conserved sequence motifs including an olf-1 site in the putative promoter region of these genes and subsequently several transcription factors were identified which did bind to these sites. However, it remained elusive if an interaction between the transcription factors and the putative promoter sites may have functional implications. Therefore, a heterologous system was employed to assess the consequence of an interaction between the putative promoters and the transcription factors. HEK 293 cells were cotransfected with a reporter gene under the control of putative mOR37 promoter regions and an expression vector based gene encoding the transcription factor. The expression rate of the reporter gene was monitored by measuring luciferase activity. It was found that the three O/E transcription factors (O/E-1, O/E-2 and O/E-4) induced significant activation of the mOR37 promoters; in addition, it was observed that the putative promoters of other OR genes were also activated, suggesting that the O/E proteins may play a general role in the regulation of OR gene expression. Mutagenesis experiments revealed that the effects of O/E proteins were dependent on the presence of an olf-1 site within the promoter region. For the transcription factor Lhx-2 it was found that not all but only promoters of distinct OR-genes were affected. For the mOR37 promoters a simultaneous action of O/E protein and Lhx-2 elicited an increase of reporter gene expression. The data indicate that the putative mOR37 promoters could drive gene expression in the presence of the crucial transcription factors in this heterologous system. In order to explore to what extent the promoter may contribute to the characteristic topographic expression pattern of the mOR37 genes in vivo, a mOR37C transgene which included the coding exon and the putative promoter, was randomly inserted into the mouse genome. Seven lines were obtained; in all lines the transgene was specifically expressed in olfactory sensory neurons (OSN). In six lines the transgene expression was restricted to the central patch of the olfactory turbinates, typical for the OR37 genes. In one line (line 7) the transgene was also

expressed in OSNs ectopically positioned outside the patch within the medial zone. It was found that the transgene was expressed in a mutually exclusive manner and from only one allele. The axons of OSNs expressing the transgene co-converged in the same glomerulus with the axons from neurons expressing the endogenous gene. In line #7 the formation of ectopic glomeruli was observed. The number of OSNs expressing the transgene varied considerably among lines; these differences were independent from the copy number of the transgene. The data indicate that the short putative promoters, most likely the conserved motifs, were sufficient to drive the OR37 gene expression in a tissue specific way and most aspects of the OR37 gene expression were mimicked by the transgene; however, considerable differences between certain lines suggested additional regulatory elements, such as a locus control region (LCR). Since regulatory elements for gene transcription, such as promoters, enhancers and LCRs, appear to be conserved across species, a comparative approach was utilized to search for the LCR-like element for the OR37 locus by sequence alignment across distantly related mammals. A segment of 270 base pairs located 137 Kb upstream of OR37 cluster I was found to be highly conserved between mouse, human, dog and opossum. It was not associated with an exon of any known gene and was highly correlated with OR37 cluster I rather than with the neighboring genes, since the flanking genes did not show syntenic conservation in the opossum genome. A homologous counterpart for this segment was found downstream of the OR37 cluster II locus; an alignment of the cluster II sequence across species identified the conservation of this counterpart. Examination for relevant motifs in this segment and comparison with the conserved H element revealed two common transcription factor binding sites, at least one of them is known to be essential for generating DNase I hypersensitive sites in the LCR of the beta globin gene locus. Further studies are required to evaluate a possible role of this conserved segment in the regulation of the OR37 gene expression.

6. Zusammenfassung

Die olfaktorischen Rezeptoren (OR) der OR37-Familie sind durch eine verlängerte dritte extrazelluläre Domäne und ihre spezifische Expression in olfaktorischen Sinneszellen (OSZ) eines kleinen Areals der zentralen Nasenturbine charakterisiert. Die kodierenden Gene dieser distinkten OR-Familie sind an 2 Loci (Cluster I und II) auf dem Chromosom 4 der Maus lokalisiert. Im Hinblick auf ein Verständnis der molekularen Grundlagen für das einzigartige topographische Expressionsmuster dieser Gene wurden in früheren Untersuchungen hochkonservierte Sequenzmotive, inklusive eine olf-1 Bindungsstelle, in ihren putativen Promotorregionen ermittelt; darüber hinaus wurden mehrere Transkriptionsfaktoren identifiziert, die von diesen Promotor-Regionen gebunden werden. Ob Interaktionen dieser Transkriptionsfaktoren und der Promotorregion funktionelle Auswirkungen auf eine Expression der Gene haben, ist noch ungeklärt. Diese Frage wurde mit Hilfe eines heterologen Systems untersucht; dazu wurde in HEK 293-Zellen ein Reporter gen unter der Kontrolle von Promotorregionen der mOR37-Gene eingeschleust und Transkriptionsfaktoren mittels spezifischer Expressionsvektoren generiert. Die Expressionsrate des Reporters wurde durch die Ermittlung der jeweiligen Luciferaseaktivität bestimmt. Dabei zeigte sich, dass drei Transkriptionsfaktoren der O/E-Familie (O/E-1, O/E-2 und O/E-4) eine signifikante Aktivierung der mOR37 Promotoren induzierten; allerdings wurden auch putative Promotoren anderer OR-Gene aktiviert. Mutagenese-Experimente haben gezeigt, dass der Einfluss der O/E-Proteine die Anwesenheit einer intakten olf-1 Bindungsstelle der Promotorregion benötigt. Diese Befunde sprechen dafür, dass O/E-Proteinen eine generelle Bedeutung für die Regulation der OR-Expression zukommen könnte. Für den Transkriptionsfaktor Lhx-2 wurde festgestellt, dass nur Promotoren von distinkten OR-Genen beeinflusst wurden. Eine gleichzeitige Aktion von O/E- und von Lhx-2-Proteinen bewirkte eine erhöhte Expressionsrate des Reporters. Diese Ergebnisse sprechen für eine Beteiligung der putativen Promotoren und der identifizierten Transkriptionsfaktoren an der Expressionskontrolle von OR37-Genen. Um herauszufinden, in welchem Maße die identifizierte putative Promotor-Region das charakteristische topographische Expressionsmuster der mOR37-Gene *in vivo* bestimmt, wurde ein Transgen von mOR37C, bestehend aus dem kodierenden Exon sowie dem putativen Promotor generiert und anschließend zufällig in das Mausgenom integriert, um zu prüfen, ob die putative

Promotor-Region der OR37-Gene alle Informationen für die zellspezifische und topographische Expression beinhaltet oder ob die Integrationsstelle im Genom entsprechenden Einfluß ausübt. In den sieben generierten Mauslinien wurde das Transgen exklusiv in OSZ exprimiert. In 6 Linien beschränkte sich die Expression des Transgens auf die typische zentrale OR37-Region des olfaktorischen Epithels. In einer Linie (Linie 7) wurde das Transgen auch ektopisch, d.h. außerhalb der zentralen Region exprimiert. In den entsprechenden OSZ wurde ausschließlich das Transgen und nur ein Allel exprimiert (monoallelisch). Die Anzahl der transgen-exprimierenden OSZ variierte zwischen den Linien beträchtlich, wobei allerdings keine Korrelation mit der Anzahl der Transgen-Kopien bestand. Die Axone der transgen-exprimierenden OSZ projizierten in denselben Glomerulus wie die Axone von Neuronen, die das endogene Gen exprimierten. In Linie 7 wurde die Bildung von ektopischen Glomeruli beobachtet. Die Ergebnisse zeigen, dass eine relativ kurze Promoter-Region offenbar ausreicht, um die gewebespezifische und regionsspezifische Expression der OR37-Gene sicherzustellen; die beträchtlichen Unterschiede in der Zellzahl deutet jedoch auf zusätzliche Regulationselemente hin. Da Regulationselemente für die Transkription von Genen, wie Promotoren, Enhancer und Locuskontrollregionen (LCR) oft speziesübergreifend konserviert sind, wurde durch Sequenzvergleiche von weitläufig verwandten Säugetierspezies versucht, solche konservierten Sequenz-Abschnitte für die OR37-Loci zu ermitteln. Dabei wurde ein Segment von 270 Basenpaaren entdeckt, das sich 137 Kb oberhalb des OR37-Clusters I befindet und zwischen Maus, Mensch, Hund und Opossum hochkonserviert ist. Das 270 bp Segment korrelierte sehr viel stärker mit dem OR37 Cluster I als mit benachbarten Genen, da die begleitenden Gene keine syntenische Konservierung im Opossumgenom aufwiesen. In der Nähe von Cluster-II des OR37-Locus wurde ein entsprechendes Element entdeckt; ein Spezies- Vergleich des Segments für Cluster II zeigte die Konservierung dieses Elementes. Untersuchungen hinsichtlich potentiell relevanter Motive in diesem Segment und Vergleiche mit dem konservierten H element haben zu Identifizierung von zwei Bindungsstellen für Transkriptionsfaktoren geführt; eine davon entspricht der essentiellen hypersensitiven DNase-I Stelle in der LCR des Betaglobin-Genlocus. Zukünftige molekulargenetische Studien sollten zur Klärung der funktionellen Bedeutung dieses konservierten DNA-Segmentes für die Regulation der Expression von OR37-Genen beitragen.

7. References

Aloni R, Olender T, and Lancet D. (2006). Ancient genomic architecture for mammalian olfactory receptor clusters. *Genome Biol.* 7:R88.

Aronow BJ, Silbiger RN, Dusing MR, Stock JL, Yager KL, Potter SS, Hutton JJ, and Wiginton DA. (1992). Functional analysis of the human adenosine deaminase gene thymic regulatory region and its ability to generate position-independent transgene expression. *Mol Cell Biol.* 12:4170-4185.

Baker JE, Kang J, Xiong N, Chen T, Cado D, and Raulet DH. (1999). A novel element upstream of the Vgamma2 gene in the murine T cell receptor gamma locus cooperates with the 3' enhancer to act as a locus control region. *J Exp Med.* 190:669-679.

Barnea G, O'Donnell S, Mancina F, Sun X, Nemes A, Mendelsohn M, and Axel R. (2004). Odorant receptors on axon termini in the brain. *Science.* 304:1468.

Bonifer C, Vidal M, Grosveld F, and Sippel AE. (1990). Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. *EMBO J.* 9:2843-2848.

Bozza T, Feinstein P, Zheng C, Mombaerts P. (2002). Odorant receptor expression defines functional units in the mouse olfactory system. *J. Neurosci.* 22:3033-3043.

Buck L, Axel R. (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* 65:175-187.

Buhler J. (2001). Efficient large-scale sequence comparison by locality-sensitive hashing. *Bioinformatics.* 17:419-28.

Bulger M, and Groudine M. (1999). Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.* 13:2465-2477.

Cao A, and Moi P. (2002). Regulation of the globin genes. *Pediatr Res.* 51:415-421.

Carter D, Chakalova L, Osborne CS, Dai YF, and Fraser P. (2002). Long-range chromatin regulatory interactions in vivo. *Nat Genet.* 32:623-626.

Caterina JJ, Ciavatta DJ, Donze D, Behringer RR, and Townes TM. (1994). Multiple elements in human beta-globin locus control region 5' HS 2 are involved in enhancer activity and position-independent, transgene expression. *Nucleic Acids Res.* 22:1006-1011.

Chess A, Simon I, Cedar H, Axel R. (1994). Allelic inactivation regulates olfactory receptor gene expression. *Cell* 78:823-834.

Cheung TH, Kwan YL, Hamady M, and Liu XD. (2006). Unraveling transcriptional control and cis-regulatory codes using the software suite GeneACT. *Genome Biol.* 7:R97.

Cole ST. (1983). Characterisation of the promoter for the LexA regulated *sulA* gene of *Escherichia coli*. *Mol Gen Genet.* 189:400-404.

Davies N, Freebody J, and Murray V. (2004). Chromatin structure at the flanking regions of the human beta-globin locus control region DNase I hypersensitive site-2: proposed nucleosome positioning by DNA-binding proteins including GATA-1. *Biochim Biophys Acta.* 1679:201-213.

Dekker J, Rippe K, Dekker M, and Kleckner N. (2002). Capturing chromosome conformation. *Science.* 295:1306-1311.

Dillon N, Trimborn T, Strouboulis J, Fraser P, and Grosveld F. (1997). The effect of distance on long-range chromatin interactions. *Mol Cell.* 1:131-139.

Ebrahimi FA, Chess A. (2000). Olfactory neurons are interdependent in maintaining axonal projections. *Curr.Biol.* 10:219-222.

Feinstein P, Mombaerts P. (2004). A contextual model for axonal sorting into glomeruli in the mouse olfactory system. *Cell* 117:817-831.

Feinstein P, Bozza T, Rodriguez I, Vassalli A, Mombaerts P. (2004). Axon Guidance of Mouse Olfactory Sensory Neurons by Odorant Receptors and the beta2 Adrenergic Receptor. *Cell* 117:833-846.

Forrester WC, Takegawa S, Papayannopoulou T, Stamatoyannopoulos G, Groudine M.(1987). Evidence for a locus activation region: the formation of developmentally

stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res.* 15:10159-10177.

Fraser P, and Grosveld F. (1998). Locus control regions, chromatin activation and transcription. *Curr Opin Cell Biol.* 10:361-365.

Godfrey PA, Malnic B, and Buck LB. (2004). The mouse olfactory receptor gene family. *Proc Natl Acad Sci U S A.* 101:2156-2161.

Grosveld F, van Assendelft GB, Greaves DR, and Kollias G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell.* 51:975-985.

Hardison R, Slightom JL, Gumucio DL, Goodman M, Stojanovic N, and Miller W. (1997). Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene.* 205:73-94.

Hagman J, Belanger C, Travis A, Turck CW, and Grossscheldl R. (1993). Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. *Genes Dev.* 7:760-773.

Hardison RC, Oeltjen J, and Miller W. (1997). Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. *Genome Res.* 7:959-966.

Hirota J, and Mombaerts P. (2004). The LIM-homeodomain protein Lhx2 is required for complete development of mouse olfactory sensory neurons. *Proc Natl Acad Sci USA* 101:8751-8755.

Hirota J, Omura M, and Mombaerts P. (2007). Differential impact of Lhx2 deficiency on expression of class I and class II odorant receptor genes in mouse. *Mol Cell Neurosci* 34:679-688.

Hobert O, and Westphal H. (2000). Functions of LIM-homeobox genes. *Trends Genet* 16:75-83.

Hoppe R, Breer H, Strotmann J. (2003a). Organization and evolutionary relatedness of OR37 olfactory receptor genes in mouse and human. *Genomics* 82:355-364.

- Hoppe R, Breer H, Strotmann J. (2006b). Promoter motifs of olfactory receptor genes expressed in distinct topographic patterns. *Genomics* 87:711-723.
- Hoppe R, Frank H, Breer H, Strotmann J. (2003b). The clustered olfactory receptor gene family 262: genomic organization, promoter elements, and interacting transcription factors. *Genome Res.* 13:2674-2685.
- Hoppe R, Lambert TD, Samollow, Breer H, and Strotmann J. (2006a). Evolution of the "OR37" subfamily of olfactory receptors: a cross-species comparison. *J Mol Evol.* 62:460-472.
- Hoppe R, Weimer M, Beck A, Breer H, Strotmann J. (2000). Sequence analyses of the olfactory receptor gene cluster mOR37 on mouse chromosome 4. *Genomics* 66:284-295.
- Huang Y, Liu DP, Wu L, Li TC, Wu M, Feng DX, and Liang CC. (2000). Proper developmental control of human globin genes reproduced by transgenic mice containing a 160-kb BAC carrying the human beta-globin locus. *Blood Cells Mol Dis.* 26:598-610.
- Im H, Grass JA, Johnson KD, Kim SL, Boyer ME, Imbalzano AN, Bieker JJ, and Bresnick EH. (2005). Chromatin domain activation via GATA-1 utilization of a small subset of dispersed GATA motifs within a broad chromosomal region. *Proc Natl Acad Sci U S A.* 102:17065-17070.
- Imai T, Suzuki M, Sakano H. (2006). Odorant receptor-derived cAMP signals direct axonal targeting. *Science* 314:657-661.
- Ishii T, Serizawa S, Kohda A, Nakatani H, Shiroishi T, Okumura K, Iwakura Y, Nagawa F, Tsuboi A, Sakano H. (2001). Monoallelic expression of the odourant receptor gene and axonal projection of olfactory sensory neurones. *Genes Cells.* 6:71-78.
- Iwema CL, Fang H, Kurtz DB, Youngentob SL, Schwob JE. (2004). Odorant receptor expression patterns are restored in lesion-recovered rat olfactory epithelium. *J Neurosci.* 24:356-369.

- Jarman AP, Wood WG, Sharpe JA, Gourdon G, Ayyub H, and Higgs DR. (1991). Characterization of the major regulatory element upstream of the human alpha-globin gene cluster. *Mol Cell Biol.* 11:4679-4689.
- Joulin V, and Richard-Foy H. (1995). A new approach to isolate genomic control regions. Application to the GATA transcription factor family. *Eur J Biochem.* 232:620-626.
- Kim A, and Dean A. (2004). Developmental stage differences in chromatin subdomains of the beta-globin locus. *Proc Natl Acad Sci U S A.* 101:7028-7033.
- Kotkow KJ, and Orkin SH. (1995). Dependence of globin gene expression in mouse erythroleukemia cells on the NF-E2 heterodimer. *Mol Cell Biol.* 15:4640-4647.
- Kubick S, Strotmann J, Andreini I, Breer H. (1997). Subfamily of olfactory receptors characterized by unique structural features and expression patterns. *J. Neurochem.* 69:465-475.
- Kudrycki K, Stein-izsak C, Behn C, Grillo M, Akeson R, and Mraolis FL. (1993). Olf-1 binding site: characterization of an olfactory neuron-specific promoter motif. *Mol Cell Biol* 13:3002-3014.
- Kudrycki KE, Buiakova O, Tarozzo G, Grillo M, Walters E, and Margolis FL. (1998). Effects of mutation of the Olf-1 motif on transgene expression in olfactory receptor neurons. *J Neurosci Res.* 52:159-172.
- Layon ME, Ackley CJ, West RJ, and Lowrey CH. (2007). Expression of GATA-1 in a non-hematopoietic cell line induces beta-globin locus control region chromatin structure remodeling and an erythroid pattern of gene expression. *J Mol Biol.* 366:737-44.
- Levai O, Breer H, Strotmann J. (2003). Subzonal organization of olfactory sensory neurons projecting to distinct glomeruli within the mouse olfactory bulb. *J. Comp Neurol.* 458:209-220.
- Lewcock JW, Reed RR. (2004). A feedback mechanism regulates monoallelic odorant receptor expression. *Proc. Natl. Acad. Sci. U.S.A.* 101:1069-74.

- Li Q, Peterson KR, and Stamatoyannopoulos G. (1998). Developmental control of epsilon- and gamma-globin genes. *Ann N Y Acad Sci.* 850:10-17.
- Li Q, Zhang M, Duan Z, Stamatoyannopoulos G. (1999). Structural analysis and mapping of DNase I hypersensitivity of HS5 of the beta-globin locus control region. *Genomics.* 61:183-93.
- Li Q, Zhou B, Powers P, Enver T, and Stamatoyannopoulos G. (1991). Primary structure of the goat beta-globin locus control region. *Genomics.* 9:488-499.
- Li QL, Peterson KR, Fang XD, and Stamatoyannopoulos G. (2002). Locus control regions. *Blood.* 100:3077-3086.
- Li Q, Han H, Ye X, Stafford M, Barkess G, and Stamatoyannopoulos G. (2004). Transcriptional potentials of the beta-like globin genes at different developmental stages in transgenic mice and hemoglobin switching. *Blood Cells Mol Dis.* 33:318-325.
- Liberg D, Sigvardsson M, and Åkerblad P. (2002). The EBF/Olf/Collier family of transcription factors: regulators of differentiation in cells originating from all three embryonal germ layers. *Mol Cell Biol* 22:8389-8397.
- Liu LR, Du ZW, Zhao HL, Liu XL, Huang XD, Shen J, Ju LM, Fang FD, and Zhang JW. (2005). T to C substitution at -175 or -173 of the gamma-globin promoter affects GATA-1 and Oct-1 binding in vitro differently but can independently reproduce the hereditary persistence of fetal hemoglobin phenotype in transgenic mice. *J Biol Chem.* 280:7452-7459.
- Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell* 126:403-413.
- Malnic B, Godfrey PA, and Buck LB. (2004). The human olfactory receptor gene family. *Proc Natl Acad Sci U S A.* 101:2584-2589.
- Malnic B, Hirono J, Sato T, Buck LB. (1999). Combinatorial receptor codes for odors. *Cell.* 96:713-723.

- Michaloski JS, Galante PA, Malnic B. (2006). Identification of potential regulatory motifs in odorant receptor genes by analysis of promoter sequences. *Genome Res.* 16:1091-1098.
- Milot E, Strouboulis J, Trimborn T, Wijgerde M, de Boer E, Langeveld A, Tan-Un K, Vergeer W, Yannoutsos N, Grosveld F, and Fraser P. (1996). Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell.* 87:105-114.
- Miyamichi K, Serizawa S, Kimura HM, Sakano H. (2005). Continuous and overlapping expression domains of odorant receptor genes in the olfactory epithelium determine the dorsal/ventral positioning of glomeruli in the olfactory bulb. *J Neurosci* 25:3586-3592.
- Molete JM, Petrykowska H, Sigg M, Miller W, and Hardison R. (2002). Functional and binding studies of HS3.2 of the beta-globin locus control region. *Gene.* 283:185-197.
- Mombaerts P, Wang F, Dulac C, Chao SK, Nemes A, Mendelsohn M, Edmondson J, Axel R. (1996). Visualizing an olfactory sensory map. *Cell* 87:675-686.
- Mombaerts P. (2004). Genes and ligands for odorant, vomeronasal and taste receptors. *Nat.Rev.Neurosci.* 5:263-278.
- Mombaerts P. (2006). Axonal wiring in the mouse olfactory system. *Annu.Rev.Cell Dev Biol* 22:713-737.
- Moon AM, and Ley TJ. (1999). Conservation of the primary structure, organization, and function of the human and mouse beta-globin locus-activating regions. *Proc Natl Acad Sci U S A.* 87:7693-7697.
- Niimura Y, and Nei M. (2005). Comparative evolutionary analysis of olfactory receptor gene clusters between humans and mice. *Gene.* 346:13-21.
- Ptashne M. (1986). Gene regulation by proteins acting nearby and at a distance. *Nature.* 322:697-701.
- Pyrski M, Xu Z, Walters E, Gilbert DJ, Jenkins NA, Copeland NG, Margolis FL. (2001). The OMP-lacZ transgene mimics the unusual expression pattern of OR-Z6, a

new odorant receptor gene on mouse chromosome 6: implication for locus-dependent gene expression. *J.Neurosci.* 21:4637-4648.

Qasba P, Reed RR. (1998). Tissue and zonal-specific expression of an olfactory receptor transgene. *J.Neurosci.* 18:227-236.

Regales L, Giraldo P, Garcia-Diaz A, Lavado A, and Montoliu L. (2003). Identification and functional validation of a 5' upstream regulatory sequence in the human tyrosinase gene homologous to the locus control region of the mouse tyrosinase gene.

Reitman M, Lee E, Westphal H, and Felsenfeld G. (1990). Site-independent expression of the chicken beta A-globin gene in transgenic mice. *Nature.* 348:749-752.

Ressler KJ, Sullivan SL, Buck LB. (1993). A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell* 73:597-609.

Ressler KJ, Sullivan SL, Buck LB. (1994). Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell* 79:1245-1255.

Rothman A, Feinstein P, Hirota J, Mombaerts P. (2005). The promoter of the mouse odorant receptor gene M71. *Mol.Cell Neurosci* 28:535-546.

Ryan TM, Sun CW, Ren J, and Townes TM. (2000). Human gamma-globin gene promoter element regulates human beta-globin gene developmental specificity. *Nucleic Acids Res.* 28:2736-2740.

Sambrook J, Frisch EF, Maniatis T. (1989). *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, and Miller W. (2000). PipMaker-a web server for aligning two genomic DNA sequences. *Genome Res.* 10:577-586

Serizawa S, Ishii T, Nakatani H, Tsuboi A, Nagawa F, Asano M, Sudo K, Sakagami J, Sakano H, Ijiri T, Matsuda Y, Suzuki M, Yamamori T, Iwakura Y, Sakano H.

(2000). Mutually exclusive expression of odorant receptor transgenes. *Nat. Neurosci.* 3:687-693.

Serizawa S, Miyamichi K, Nakatani H, Suzuki M, Saito M, Yoshihara Y, Sakano H. (2003). Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science.* 302:2088-2094.

Shetty RS, Bose SC, Nickell MD, McIntyre JC, Hardin DH, Harris AM, and McClintock TS. (2005). Transcriptional changes during neuronal death and replacement in the olfactory epithelium. *Mol Cell Neurosci.* 30:90-107.

Shykind BM, Rohani SC, O'Donnell S, Nemes A, Mendelsohn M, Sun Y, Axel R. Barnea, G. (2004). Gene switching and the stability of odorant receptor gene choice. *Cell.* 117:801-815.

Smit A, and Green P. (1999). *RepeatMasker* at <http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

Strotmann J, Breer H. (2006). Formation of glomerular maps in the olfactory system. *Semin Cell Dev Biol* 17:402-410.

Strotmann J, Conzelmann S, Beck A, Feinstein P, Breer H, Mombaerts P. (2000). Local permutations in the glomerular array of the mouse olfactory bulb. *J. Neurosci.* 20:6927-6938.

Strotmann J, Hoppe R, Conzelmann S, Feinstein P, Mombaerts P, Breer H. (1999). Small subfamily of olfactory receptor genes: structural features, expression pattern and genomic organization. *Gene* 236:281-291.

Strotmann J, Levai O, Fleischer J, Schwarzenbacher K, Breer H. (2004). Olfactory receptor proteins in axonal processes of chemosensory neurons. *J Neurosci.* 24:7754-7761.

Strotmann J, Wanner I, Helfrich T, Beck A, Breer H. (1994a). Rostro-caudal patterning of receptor-expressing olfactory neurones in the rat nasal cavity. *Cell Tissue Res.* 278:11-20.

Strotmann J, Wanner I, Helfrich T, Beck A, Meinken C, Kubick S, Breer H. (1994b). Olfactory neurones expressing distinct odorant receptor subtypes are spatially segregated in the nasal neuroepithelium. *Cell Tissue Res.* 276:429-438.

Strotmann J, Wanner I, Helfrich T, and Breer H. (1995). Receptor expression in olfactory neurons during rat development: in situ hybridization studies. *Eur J Neurosci.* 7:492-500.

Strotmann J, Wanner I, Krieger J, Raming K, Breer H. (1992). Expression of odorant receptors in spatially restricted subsets of chemosensory neurones. *NeuroReport* 3:1053-1056.

Tolhuis B, Palstra RJ, Splinter E, Grosveld F, and de Laat W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell.* 10:1453-1465.

Touhara K, Sengoku S, Inaki K, Tsuboi A, Hirono J, Sato T, Sakano H, Haga T. (1999). Functional identification and reconstitution of an odorant receptor in single olfactory neurons. *Proc.Natl.Acad.Sci.U.S.A* 96:4040-4045.

Tsai RY, and Reed RR. (1997). Cloning and functional characterization of Roaz, a zinc finger protein that interacts with O/E-1 to regulate gene expression: implications for olfactory neuronal development. *J Neurosci.* 17:4159-4169.

Tuan D, Solomon W, Li Q, and London IM. (1985). The "beta-like-globin" gene domain in human erythroid cells. *Proc Natl Acad Sci U S A.* 82:6384-6388.

Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker J, and Bolbel GA. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell.* 17:453-462.

Vassalli A, Rothman A, Feinstein P, Zapotocky M, Mombaerts P. (2002). Minigenes impart odorant receptor-specific axon guidance in the olfactory bulb. *Neuron* 35:681-696.

Vassar R, Chao SK, Sitcheran R, Nunez JM, Vosshall LB, Axel R. (1994). Topographic organization of sensory projections to the olfactory bulb. *Cell* 79:981-991.

- Vassar R, Ngai J, Axel R. (1993). Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. *Cell* 74:309-318.
- Wang F, Nemes A, Mendelsohn M, Axel R. (1998). Odorant receptors govern the formation of a precise topographic map. *Cell* 93:47-60.
- Wang MM, and Reed RR. (1993). Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. *Nature*. 362:764-767.
- Wang MM, and Reed RR. (1993). Molecular mechanisms of olfactory neuronal gene regulation. *Ciba Found Symp.* 179:68-72.
- Wang MM, Tsai RY, Schrader KA, and Reed RR. (1993). Genes encoding components of the olfactory signal transduction cascade contain a DNA binding site that may direct neuronal expression. *Mol Cell Biol.* 13:5805-5813.
- Wang SS, Betz AG, and Reed RR. (2002). Cloning of a novel Olf-1/EBF-like gene, O/E-4, by degenerate oligo-based direct selection. *Mol Cell Neurosci.* 20:404-414.
- Wang SS, Lewcock JW, Feinstein P, Mombaerts P, and Reed RR. (2004). Genetic disruptions of O/E2 and O/E3 genes reveal involvement in olfactory receptor neuron projection. *Development* 131:1377-1388.
- Wang SS, Tsai RY, and Reed RR. (1997). The characterization of the Olf-1/EBF-like HLH transcription factor family: implications in olfactory gene regulation and neuronal development. *J Neurosci* 17:4149-4458.
- Young JM, Friedman C, Williams EM, Ross JA, Tonnes-Priddy L, and Trask BJ. (2002). Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Hum Mol Genet.* 11:535-46.
- Young JM, Trask BJ. (2002). The sense of smell: genomics of vertebrate odorant receptors. *Hum.Mol.Genet.* 11:1153-1160.
- Zhang X, and Firestein S. (2002). The olfactory receptor gene superfamily of the mouse. *Nat Neurosci.* 5:124-133.

8. Abbreviations

ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
Bidest	Twice distilled water
bp	Base pairs
Ca²⁺	Calcium (divalent)
CNG	Cyclic nucleotide gated channel
cDNA	Complementary DNA
DIG	Digoxigenin
DNA	Deoxyribonucleic acid
DNase	Neoxyribonuclease
DTT	Dithiothreitol
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
dGTP	Deoxyguanine triphosphate
dTTP	Deoxythymidine triphosphate
dNTP	Deoxynucleotide triphosphate
E.coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EGTA	ethylene glycol bis(2-aminoethyl ether)-N,N,N',N'-tetraacetic acid
EtBr	Ethidium bromide
EtOH	Ethanol
<i>et al</i>	Others
FBS	Fetal Bovine Serum
Fig	Figure
FISH	Fluorescence in situ hybridization
GFP	Green fluorescent protein
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	Isopropylthio- β -galactosidase
Kb	Kilobase

KV	Kilo volts
LB	Luria Bertani
LCR	Locus control region
mg	Milligram
ml	Millilitre
mM	Millimolar
mm	Millimeter
mRNA	Messenger ribonucleic acid
μF	Microfarad
μg	microgram
μl	microlitre
μM	Micromolar
μm	Micrometer
NCBI	National Center for Biotechnology Information
ng	Nanogram
nM	Nanomolar
OB	Olfactory bulb
OE	Olfactory epithelium
OMP	Olfactory marker protein
OSN	Olfactory sensory neurons
PBS	phosphate buffered saline
PCR	Polymerase chain reaction
pH	Negative log of H ⁺ ion concentration
RNA	Ribonucleic acid
RNase	Ribonuclease
rmp	Rotation per minute
SDS	Sodium dodecyl sulfate
SSC	Sodium chloride sodium citrate
<i>Taq</i>	<i>Thermus aquaticus</i>
TAE	Tris-acitrate-EDTA buffer
TBE	Tris-borate-EDTA buffer
TF	Transcription factor
3C	Chromosome conformation capturing

TSS	Transcription start site
Tween 20	Polyoxyethylene sorbitan monolaurate
UV	Ultra violet
Vol	Volume
X-gal	5-Bromo-4chloro-3-indolyl-D-galactosidase

Curriculum Vitae

Name Zhang Yong-quan
Date of Birth 21 June 1975
Place of Birth Shandong, China
State of marriage Married

Education and Profession

1993-1997 **Bachelor of Science** (Department of Biology, Shandong Normal University, China)
1997-1999 **Teacher for Biology** in Jinxiang high school, China
1999-2003 **Master of Science** (Shanghai Institute of Physiology, Chinese Academy of Sciences, China)
2003-2007 **PhD** (Institute of Physiology, University of Hohenheim, Germany)

Research publication

1. Yong-Quan Zhang , Yuan-Cong Zhou and Guo-Guang Shen. (2002) beta-Agkistrodotoxin inhibition of Ca²⁺-dependent release of glutamate, aspartate, glycine and gamma-aminobutyric acid from cerebrocortical synaptosomes following its binding to synaptic membrane. *Neuroreport* 13: 2313-2317.
2. Yong-Quan Zhang, Heinz Breer and Joerg Strotmann. (2007) Promotor elements governing the clustered expression pattern of odorant receptor genes. *Mol. Cell. Neurosci.* **Accepted**.

Acknowledgement

First of all I would like to present my gratitude to Professor Dr. Heinz Breer for providing the opportunity to carry out the research for my Ph.D thesis in Institute of Physiology, Hohenheim University, and for guiding me to the scientific realm with his kind assistance and his wise instructions. Due to the support from Professor Dr. Heinz Breer my family and I are capable of having a happy and peaceful life in Germany. I thank Professor Dr. Heinz Breer for all his help also on behalf of my family.

Secondly I would like to thank Dr. Jörg Strotmann for his supervision on reading scientific papers, designing and performing the experiments, analysing the results and improving the research strategies. Many thanks for his patient discussion, constructive advice and his kind help with my daily life.

I thank my house owners Dr. Abdramane Tékété and Frau Margit Tékété for their gentle help during the period my wife was pregnant and my daughter was born and for their kind assistance for our daily life in Germany.

I thank Frau Christina Beck, Frau Eva Schmidt and Frau Gudrun Stickler for all their aids to my life and study in Germany.

I want to thank Roger Frei for his support and all his help especially in the first year after I came here. For improving my spoken English I would like to give my thanks to Tom Lambert and Bob.

I thank Jürgen Krieger, Jörg Fleischer, Reiner, Ewald, Thomas, Karin, Georg, Nicole, Heiko, Marion, Miriam, Marcus, Maike, Carina, Kerstin, Gesa and other colleagues for their help with the lab life and for enjoying the coffee together in the late afternoon.

I thank my wife and my daughter who make me realise the responsibility for a family and promote me to continue the pursuit for a happy life. Their love is always my motivation to insist on whenever I am confronted with some trouble.

Special thanks will be given to my parents from whom I obtained my life as well as the courage, bravery and strength. They always stand by me no matter I succeed or fail.

This thesis is heartedly dedicated to my grandfather and grandmother who brought me up from a child to a boy with their whole love.