**Dissertation**

**Zur Erlangung des Grades**

**Doktor der Naturwissenschaften**

**Am Fachbereich Biologie**

**Der Johannes Gutenberg-Universität Mainz**

**Ying Yue**

**geb. am 23-06-1974 in Shanghai, China**

**Mainz, 2005**

**Tag der mündlichen Prüfung: 06-06-2005**

## Abbreviations

| | |
|---|---|
| A | Adenine |
| AZA | 5-Azacytidine |
| BAC | Bacterial artificial chromosome |
| BOS | Break of synteny |
| bp | Base pair |
| C | Cytosine |
| cDNA | Complementary deoxyribonucleid acid |
| CGE | *Callithrix geoffroyi* |
| CJA | *Callithrix jacchus* |
| dATP | Deoxyadenosine triphosphate |
| DAPI | 4',6-diaminodino-2-phenylindole |
| dC | Deoxycytosine |
| dCTP | Deoxycytidine triphosphate |
| dGTP | Deoxyguanosine triphosphate |
| DIG | Digoxigenin |
| DMEM | Dulbecco's modified eagle medium |
| DMSO | Dimethylsuphoxide |
| DNA | Deoxyribonucleid acid |
| DNase | Deoxyribonuclease |
| dNTP | Deoxynucleoside triphosphate |
| DOP PCR | Degenerate oligonucleotide primed polymerase chain reaction |
| DTT | Dithiothreitol |
| dTTP | Deoxythymidine triphosphate |
| EDTA | Ethylenediaminotetracetic acid |
| EST | Expressed sequence tag |
| FBS | Fetal bovine serum |
| FISH | Fluorescence in situ hybridization |
| G | Guanine |
| GGA | *Gallus gallus* |
| GGO | *Gorilla gorilla* |
| GSP | Gene specific primer |
| HAT | Hypoxanthine/aminopterin/thymidine (medium) |
| hr | Hour |
| HSA | *Homo sapiens* |

| | |
|---|---|
| HSY | *Hylobates syndactylus* |
| IMAGE | International molecular analysis of genomes and their expression (cDNA clones) |
| kb | Kilobase (1000 bp) |
| LB | Luria Bertani (medium) |
| LINE | Long interspersed nuclear element |
| LTR | Long terminal repeat |
| M | Molar |
| Mb | Megabase |
| Min | Minute |
| mM | Milimolar |
| MMU | *Mus musculus* |
| MML | *Macaca Mulatta* |
| mRNA | Messenger ribonucleic acid |
| mya | Million years ago |
| nm | Nanometer |
| N-terminal | Amino-terminal |
| $OD_{260}$ | Optical density at 260 nm |
| $OD_{280}$ | Optical density at 280 nm |
| oligo | Oligonucleotide |
| OR | Olfactory receptor |
| PAC | P1-derived artificial chromosome |
| PTR | *Pan troglodytes* |
| PCR | Polymerase chain reaction |
| PPY | *Pongo pygmaeus* |
| RACE | Rapid amplification of cDNA ends |
| RNA | Ribonucleid acid |
| RNase | Ribonuclease |
| RNO | *Rattus norvegicus* |
| ROBO | Roundabout |
| rpm | Revolutions per minute |
| RT | Reverse transcriptase |
| RT-PCR | Reverse transcriptase PCR |
| sec | Second |
| SINE | Short interspersed nuclear element |
| SSC | Saline sodium citrate |

| | |
|---|---|
| STS | Sequence tagged site |
| T | Thymine |
| *Taq* | *Thermus aquaticus* (DNA polymerase) |
| TAE | Tris Acetate EDTA buffer |
| TE | Tris/EDTA (buffer) |
| Tris | Tris(hidroximethyl) aminomethane |
| UTR | Untranslated region |
| X-gal | 5-bromo-4-chloro-3-indolyl-β-D-galactoside |
| YAC | Yeast artificial chromosome |

**Amino Acid codes**

| ***Amino Acid*** | ***Designations*** | |
|---|---|---|
| Alanine | A | Ala |
| Arginine | R | Arg |
| Asparagine | N | Asn |
| Aspartate (= Aspartic Acid) | D | Asp |
| Aspartate or Asparagine | B | Asx |
| Cysteine | C | Cys |
| Glutamine | Q | Gln |
| Glutamate (= Glutamic Acid) | E | Glu |
| Glutamate or Glutamine | Z | Glx |
| Glycine | G | Gly |
| Histidine | H | His |
| Isoleucine | I | Ile |
| Leucine | L | Leu |
| Lysine | K | Lys |
| Methionine | M | Met |
| Phenylalanine | F | Phe |
| Proline | P | Pro |
| Serine | S | Ser |
| Threonine | T | Thr |
| Tryptophan | W | Trp |
| Tyrosine | Y | Tyr |
| Valine | V | Val |

# Contents

# Abstract

Comparative fluorescence in situ hybridization (FISH) mapping revealed four large DNA segments which have been conserved in their entirety between human chromosome 3 and Bornean orangutan chromosome 2 as well as three evolutionary breakpoints which distinguish between the human and Bornean orangutan chromosome forms. Examination of the structural and functional features of evolutionary breakpoints provides new insights into the possible effects of evolutionary rearrangements on genome function and the relationship between human chromosome pathology and evolution. FISH of human BAC clones which were assesssed in human genomic sequence to primate chromosomes, combined with precise breakpoint localizations by polymerase chain reaction (PCR) analysis of flow-sorted chromosomes and *in silico* analysis, were used to characterize the evolutionary breakpoints. None of the three breakpoints studied disrupts a validated gene(s), however they are all associated with segmental duplications. At least eleven DNA segments (>50 kb; >90% similarity) paralogous to the 3q21.3 breakpoint contig were identified in human genome, with one paralogous segment in the 3p12.3 evolutionary breakpoint contig. The two breakpoint contigs at 3q21.3 and 3p12.3 as well as with nine paralogous segments contain olfactors receptor (OR) genes of the 7E subfamily. 7E OR segmental duplications are involved in both human genomic disorders and evolutionary chromosome rearrangements. Evidently, similar mechanisms drive human chromosome pathology and evolution. The 3q21.3 breakpoint region was reused for three independent chromosomal rearrangements during primate evolution. All three evolutionary breakpoint regions coincide with breaks of chromosomal synteny in the mouse, rat and chicken genomes. These data support a nonrandom breakage model of genome evolution. The human 3p12.3- and 3q21.3-syntenic regions are associated with a partial loss of duplicated sequences in the orangutan lineage and a genomic DNA insertion in the lineage leading to humans and African great apes, suggesting that the underlying rearrangements were not simple “breakage and reunion” events. A new fetal brain specific isoform of the human *ROBO2* gene, which lies closely adjacent to the evolutionary breakpoint at human 3p12.3, was identified. The 5’ end including the first two exons of this isoform are highly duplicated in the siamang gibbon and orangutan genomes and also in the pericentromeric regions of human chromosome 20 and 22. It is possible that these duplications confer a brain-specific expression pattern on newly formed chimeric transcripts.

# 1.1. Comparative analysis of mammalian genome evolution

## 1.1.1. Approaches towards understanding genome evolution

Traditionally, three approaches have been used to analyze genome organization: cytogenetic mapping, genetic linkage mapping and physical mapping. Comparisons of G-banding chromosome patterns were first used to infer homologies of whole chromosomes or subregions between species and even across mammalian orders. Banding techniques allowed a more detailed insight into primate chromosomal evolution over 50 or more million years. It soon became clear that human and great apes may only differ by few changes in chromosome morphology (Turleau *et al.*, 1972; Dutrillaux 1979; Seuánez, 1979; Yunis and Prakash, 1982; Clemente *et al.*, 1990). Subsequent gene mapping using somatic cell hybrids confirmed that large segments of mammalian genomes are remarkably conserved.

Chromosome painting (or "Zoo" fluorescence *in situ* hybridization, Zoo-FISH) has been applied to establish chromosomal homologies among mammalian orders and in certain cases (for example, the X chromosome) between placental and marsupial mammals. In this procedure, pools of DNA sequences from a particular chromosome of a given species are amplified and fluorescently labeled. The resulting "painting probe" is then hybridized to metaphase chromosomes of a different species. This method has been applied to over 40 mammalian species from nine placental orders, in particular human painting probes have been hybridized to many different primates (reviewed by Murphy *et al.*, 2001; Jauch *et al.*, 1992).

Recently, two multicolor FISH strategies for a "bar-code"-based differentiation of the karyotypes were applied in a comparative molecular cytogenetic study of higher primates (Müller and Wienberg, 2001). In cross-species color banding (also termed Rx-FISH), the painting probes are derived from gibbon chromosomes. Since gibbon chromosomes are highly rearranged by translocations, this method could delineate up to 100 different chromosome segments in the human karyotype (Müller *et al.*, 1998). Another bar-code approach, multiplex FISH, makes use of fragmented chromosomes in human/rodent somatic cell hybrids (Müller *et al.*, 1997). Since *Alu* repeats are confined to primate DNA, *Alu*-PCR then can be used to generate human painting probes from these cell lines. The "chromosome bar code" patterns from both approaches were analyzed to obtain an overview of chromosomal rearrangements during higher primate evolution. Resulting comparative maps identified numerous intra-chromosomal rearrangements. This allowed one to resolve contradicting chromosome banding interpretations and to propose an ancestral hominoid karyotype

(Müller and Wienberg, 2001).

Although chromosome painting approaches are powerful for identifying conserved syntenies, they are limited by their inability to detect changes in gene order within conserved segments and small rearrangements or inversions within chromosome segments. Greater precision can be obtained by using subregional probes derived from chromosome microdissection or large-insert clones such as yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), and cosmids.

Even greater precision in identifying conserved segments in which the gene order is known can be achieved through gene mapping approaches using genetic linkage or radiation hybrids. Genetic linkage maps are best suited to order polymorphic microsatellite markers, but are much less efficient for developing comparative gene maps given the limited degree of coding locus polymorphism observed within most intraspecies crosses (O'Brien, 1991). In radiation hybrid (RH) mapping, the chromosomes of the species of interest are fragmented using a lethal dose of radiation on cultured cells; the cells are then fused to a thymidine-kinase-deficient recipient hamster cell line and hybrids are selected on HAT medium, on which only cells with active thymidine kinase (hybrid cells) can grow. DNA from 90-100 radiation hybrid cell clones (each containing a different random assortment of chromosome fragments from the donor species) are screened for genetic markers from that species, and the markers are then ordered using maximum-likelihood ordering strategies. When several hundred coding genes have been mapped in a donor species, their positions can be compared with those of homologous loci from the gene dense human and mouse RH maps. Whole-genome radiation hybrid gene maps have been constructed for rat, dog, cat, cow, pig and zebrafish (review by Murphy *et al.,* 2001). Recently, a genome-wide high-resolution scanning method called HAPPY (haploid equivalents of DNA and the polymerase chain reaction) mapping (Dear and Cook, 1989) was used to map markers along human chromosome 2 and great ape homologs in fine detail (de Pontbriand *et al.,* 2002). HAPPY mapping is an acellular adaptation of the radiation hybrid method with even higher resolution. Minute amounts of DNA are distributed in the wells of a 96-well microtiter plate so that each of the markers to be tested is either absent or present in a particular well. Under these conditions, each well corresponds to an RH cell line, because physically linked markers tend to be distributed in the same subset of wells, whereas unlinked markers will be randomly distributed. The order of the markers and the distance between them are then deduced from their co-segregation frequencies and their pairwise lod scores.

## 1.1.2. Rates of mammalian genome evolution and phylogeny

By comparing conserved syntenies revealed by gene maps and chromosome painting, very different rates of genome rearrangements have been observed among species (reviewed by O'Brien *et al.,* 1999). In general, the mammalian genome shows a remarkable degree of evolutionary conservation. In primates, only a handful of differences are apparent between the genomes of humans, great apes, Old and New World primates, and lemurs. On average, only one to two chromosomal exchanges occur every 10 million years among eutherian mammals. This slow rate explains why multiple chromosomes or chromosome arms are preserved as intact units despite the long divergence times between different mammalian orders. This slow rate of exchange applies to species from several orders such as humans, cat, mink, ferret, and dolphin. Dramatic exceptions to the slow rate of genome evolution are observed in nearly every mammalian order with a 3- to 10-fold increased rate of rearrangements. Among primates, gibbons have genomes that are rearranged three to four times more extensively than those of humans or great apes (Jauch *et al.,* 1992; Koehler *et al.,* 1995a & b). The same is true for certain New World primates (owl and spider monkeys) and lemurs (black lemur) (Moreschalchi *et al.,* 1997; Müller *et al.,* 1997; Consigliere *et al.,* 1996). Rodent species (mouse and rat) show very high rates of genome rearrangement (roughly one rearrangement per million years) relative to other mammal species (Stanyon *et al.,* 1999).

Associations of genes or segments syntenic to regions on two or more chromosomes of humans – the standard mammalian reference species – within a single chromosome of two or more other species can be used as phylogenetic character to aid understanding of phylogenetic relationships within a given taxonomic group (Nash *et al.,* 1998). When the same association is found in two related species, the association was probably inherited from a common ancestor of the two species. Because chromosomal breaks are probably sufficiently rare, the disruption of the associations found in some species but not in others are also phylogenomic markers. Table 1.1 summarizes 19 combinations of human chromosome segments which are found to be associated in various other species by chromosome painting (from Murphy *et al.,* 2001). Chromosome association studies support most orders of mammals and superordinal clades identified by previous phylogenetic studies. For example, the order Carnivora retains two specific chromosome-segment associations not found in any other mammalian species: 2p+20 and 3+19.

### 1.1.3. The ancestral eutherian karyotype

The form and content of the ancestral genome of all placental mammals can be derived from chromosome association data. The number of hybridization signals found for each human chromosome and its distribution in extant orders provide additional information on the ancestral genome organization of placental mammals. On the basis of their widespread occurrence in several mammalian orders (Table 1.1), the following associations homologous to human chromosomes or chromosome segments seem to represent the ancestral mammalian genome: 3+21, 4+8p+4, 7q+16p, 12p-q+22qt, 12qt+22q, 14+15 and 16q+19q. These syntenic associations are also present in outgroups such as chicken and zebrafish (Burt *et al.,* 1999 and Postlethwait *et al.,* 2000). Eleven human chromosomes, 5, 6, 9, 11, 13-15, 17, 18, X, and Y, are conserved as blocks, either as single chromosomes or associated with other chromosome segments, in different species. Nine human chromosomes, 1, 2, 7, 8, 10, 12, 16, 19, and 22, are found in, at most, two fragments. The ancestral genome of placental mammals probably had a diploid number (2n) of about 50 (Figure 1.1). Three fissions and three fusions are necessary to derive the ancestral primate karyotype from the ancestral placental mammalian karyotype, whereas only four fusions are needed from the ancestral placental mammalian karyotype to the ancestral carnivore karyotype. The diploid numbers and chromosome associations present in the reconstructed ancestral genome may be slightly modified when mapping studies will be performed in more eutherian orders. Currently, gene-mapping data are too sparse in marsupials and monotremes to identify large regions of conserved synteny with eutherian autosomes (Toder *et al.,* 1998). All species of placental mammals examined by Zoo-FISH or gene mapping are limited to just two of four superordinal clades of placental mammals: Laurasiatheria and Euarchontoglires. The two putative basal clades of mammals, Afrotheria and Xenarthra, are unrepresented. A comprehensive reconstruction of the evolution of placental mammalian genomes can only be achieved when these species are examined.

Table 1.1. Human chromosome segment associations observed in other placental mammals by Zoo-FISH or gene mapping. (adapted from Murphy *et al.*, 2001)

| Human chromosome associations | Primates | | | | Scandentia | Carnivora | | | Perrisodactyla | Cetartiodactyla | | | Chiroptera | Eulipotyphia | Lagomorpha | Rodentia | | Chicken | Zebrafish |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Chimp | Macaque | Capuchin | Lemur | Tree Shrew | Cat | Seal | Ferret | Horse | Pig | Dophin | Cow | Bat | Shrew | Rabbit | Rat | Mouse | | |
| **12+22a** | | | | 10 | 9 | B4 | m3 | 9 | 1/6/26 | 5 | 8 | 5 | 3 | | | 7 | 10&15 | 1 | |
| **12+22b** | | | | 10 | 27 | D3 | | 14 | 8 | 14 | 9 | 17 | 8 | | 21 | 12 | 5 | | |
| **16q+19q** | 7 | 6 | 3 | 5 | E2 | m9 | 16 | | 6 | 17 | 18 | 13 | hn | 5 | 19 | 8 | | LG7 | |
| **14+15** | | 9 | 1 | 11 | B3 | sm4 | 8 | 1 | 1&7 | 3 | 21 | 1 | ik | 17 | | | 5-S | | |
| **3+21** | | | | 7 | C2 | sm1 | 1 | | 13 | 5 | 1 | 3 | af | 14 | 11 | 10 | 1 | LG3 | |
| **7q+16p** | | | | | E3 | | | | 3 | 13 | 29 | 4 | | 6 | | | | | |
| **4+8p** | | | | | B1 | sm2 | 5 | | | | | | | 2 | | | | | |
| **10+16** | | 5 | | 1 | | | | | | | | | | | | | | | |
| **18+22** | | | | | D3 | m7 | 14 | 8 | | 9 | | | | | | | | | |
| **12q+22+10p** | | | | | B4 | m3 | 9 | 1-S | 14-S | | | | | | | | | | |
| **3+19** | | | | | A2 | sm1 | 1 | | | | | | | | | | | | |
| **2+20** | | | | | A3 | m5 | 11 | | | | | | | | | | | | |
| **11+19** | | | | | | | | 7 | 2 | | | | | | | | | | |
| **5+19** | | | | | | | | | 2 | | | 2 | | | | | | | |
| **8+13** | | | | | | | | | | | | | 8 | bc | | | | | |
| **3+8** | | | | | | | | | | | | | 7 | af | | | | | |
| **9+11** | | | | | | | | | | | | | | | 1 | 1&3 | 2&19 | | |
| **5+8** | | | | | | | | | | | | | | | 3 | 2 | 15 | | |
| **2+7** | | | | | | | | | | | | | | | 7 | 4/6/14 | 6/11/12 | | |

Numbers represent the chromosome numbers in the indicated species that displays the human chromosome association shown in the left column. S, synteny uncertain because the homology was not found in reciprocal chromosome painting.

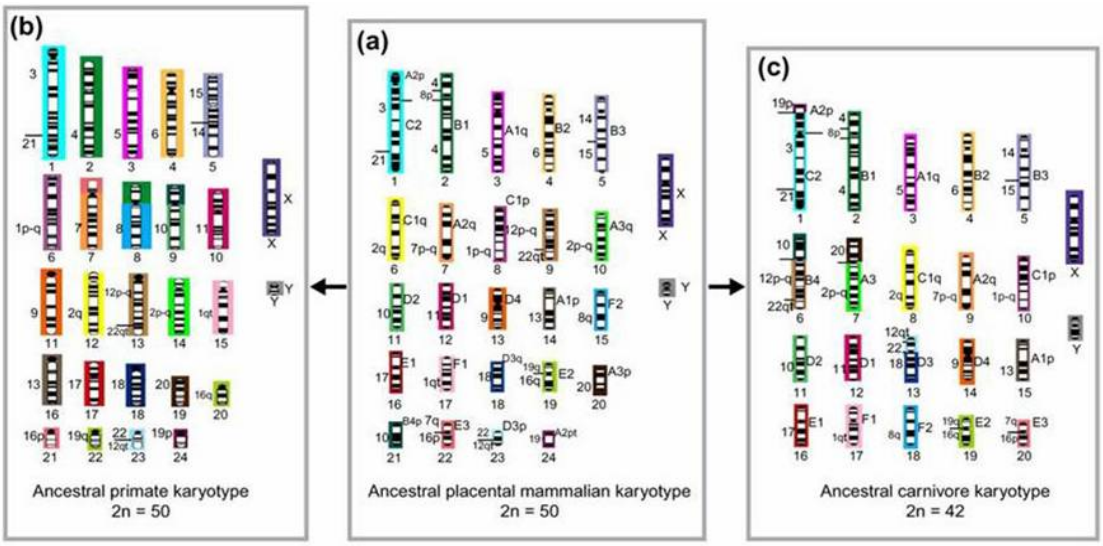



Figure 1.1. Hypothetical karyotypes of the ancestors of (a) placental mammals, (b) primates, and (c) carnivores. Numbers at the left and at the right of each chromosome represent homologous regions in the human and cat genomes respectively. The different colors represent the chromosomes in the ancestral mammalian karyotype. (From Murphy *et al.*, 2001)

## 1.2. Molecular cytogenetic analyses of the hominoid karyotype

### 1.2.1. Reconstruction of genomic rearrangements in hominoids by chromosome painting

Interspecies chromosome *in situ* suppression (CISS) hybridization of all human chromosome-specific DNA libraries has been used to establish homologies between the human, great ape (chimpanzee, gorilla, and orangutan) and gibbon karyotypes (Jauch *et al.*, 1992). Table 1.2 summarizes the chromosome homologies among hominoids as revealed by Zoo-FISH. Besides the fusion leading to human chromosome 2, only one translocation was found in great ape karyotypes between the gorilla homologs of human chromosomes 5 and 17. In contrast to this stability, many translocations were detected in gibbon karyotypes.

Table 1.2. Chromosome homologies between human and hominoids (chimpanzee, gorilla, orangutan, and gibbon) revealed by chromosome painting with human chromosome-specific DNA libraries. (From Jauch et al., 1992)

| Human chromosome-specific DNA library | Homologous chromosomes or chromosome regions | | | |
|---|---|---|---|---|
| | Chimpanzee | Gorilla | Orangutan | Gibbon* |
| Chromosome 1 | *PTR* 1 | *GGO* 1 | *PPY* 1 | *HLA* 5q; 7q; 9p; 19 |
| 2 | *PTR* 12; 13 | *GGO* 12; 11 | *PPY* 12; 11 | *HLA* 1p; 10p; 12p; 16p; 16q |
| 3 | *PTR* 2 | *GGO* 2 | *PPY* 2 | *HLA* 4p; 7q; 10q; 12q |
| 4 | *PTR* 3 | *GGO* 3 | *PPY* 3 | *HLA* 2q; 3q; 18q |
| 5 | *PTR* 4 | *GGO* 4q; 19p; 19q | *PPY* 4 | *HLA* 6p; 6q; 8p or 8q†; 18p |
| 6 | *PTR* 5 | *GGO* 5 | *PPY* 5 | *HLA* 3p; 3q; 20 |
| 7 | *PTR* 6 | *GGO* 6 | *PPY* 10 | *HLA* 1p; 1q |
| 8 | *PTR* 7 | *GGO* 7 | *PPY* 6 | *HLA* 7q; 9p; 9q |
| 9 | *PTR* 11 | *GGO* 13 | *PPY* 13 | *HLA* 8q; 13q |
| 10 | *PTR* 8 | *GGO* 8 | *PPY* 7 | *HLA* 2p; 2q; 3q |
| 11 | *PTR* 9 | *GGO* 9 | *PPY* 8 | *HLA* 11 |
| 12 | *PTR* 10 | *GGO* 10 | *PPY* 9 | *HLA* 7p; 7q; 10p; 10q; 14p; 14q |
| 13 | *PTR* 14 | *GGO* 14 | *PPY* 14 | *HLA* 4q |
| 14 | *PTR* 15 | *GGO* 18 | *PPY* 15 | *HLA* 17 |
| 15 | *PTR* 16 | *GGO* 15 | *PPY* 16 | *HLA* 15p; 15q |
| 16 | *PTR* 18 | *GGO* 17 | *PPY* 18 | *HLA* 6p; 6q; 8p |
| 17 | *PTR* 19 | *GGO* 4p; 4q; 19q | *PPY* 19 | *HLA* 8q; 13p; 16p; 16q |
| 18 | *PTR* 17 | *GGO* 16 | *PPY* 17 | *HLA* 5p |
| 19 | *PTR* 20 | *GGO* 20 | *PPY* 20 | *HLA* 10p; 14p; 14q; 16q |
| 20 | *PTR* 21 | *GGO* 21 | *PPY* 21 | *HLA* 21 |
| 21 | *PTR* 22 | *GGO* 22 | *PPY* 22 | *HLA* 15q |
| 22 | *PTR* 23 | *GGO* 23 | *PPY* 23 | *HLA* 8p; 8q |
| X | *PTR* X | *GGO* X | *PPY* X | *HLA* X |
| Y | *PTR* Y | *GGO* Y | *PPY* Y | *HLA* Y |

* *Hylobates* sp. (2n=44). † Gibbon chromosome 8 is polymorphic.

### 1.2.2. “Bar-coding” patterns of hominoid karyotypes

Rx-FISH and multiplex FISH have been used to establish the “chromosome bar code” patterns in order to reconstruct the ancestral genome organization of hominoids using the macaque as outgroup species (Müller and Wienberg, 2001). 160 individual homologous chromosomal sub-regions could be assigned to great ape and macaque G-banded ideograms (Figure 1.2, from Müller and Wienberg, 2001). Given an ancestral chromosome number of 2n = 48 for all hominoids (Dutrillaux 1979), 21 ancestral forms of the 25 different chromosomes were proposed (summarized in table 1.3). The orangutan conserved 20 ancestral great ape chromosome forms, of which 13 chromosomes are shared with the chimpanzee, 13 with humans, and 10 with gorilla (Figure 1.3). Four chromosomes were not included in this ancestral karyotype model because of extensive rearrangements close to the centromeres (chromosome IIp, IV, IX and ***Y*** chromosome), where no informative bars were available. Fine mapping of these chromosomes showed that the pericentromeric segments underwent multiple changes (Wienberg *et al.,* 1994; Nickerson and Nelson, 1998; Archidiacono *et al.,* 1998). For example, all great ape homologs of human chromosome 4 differ from the human form by distinct pericentric inversions (Marzella *et al.,* 2000).

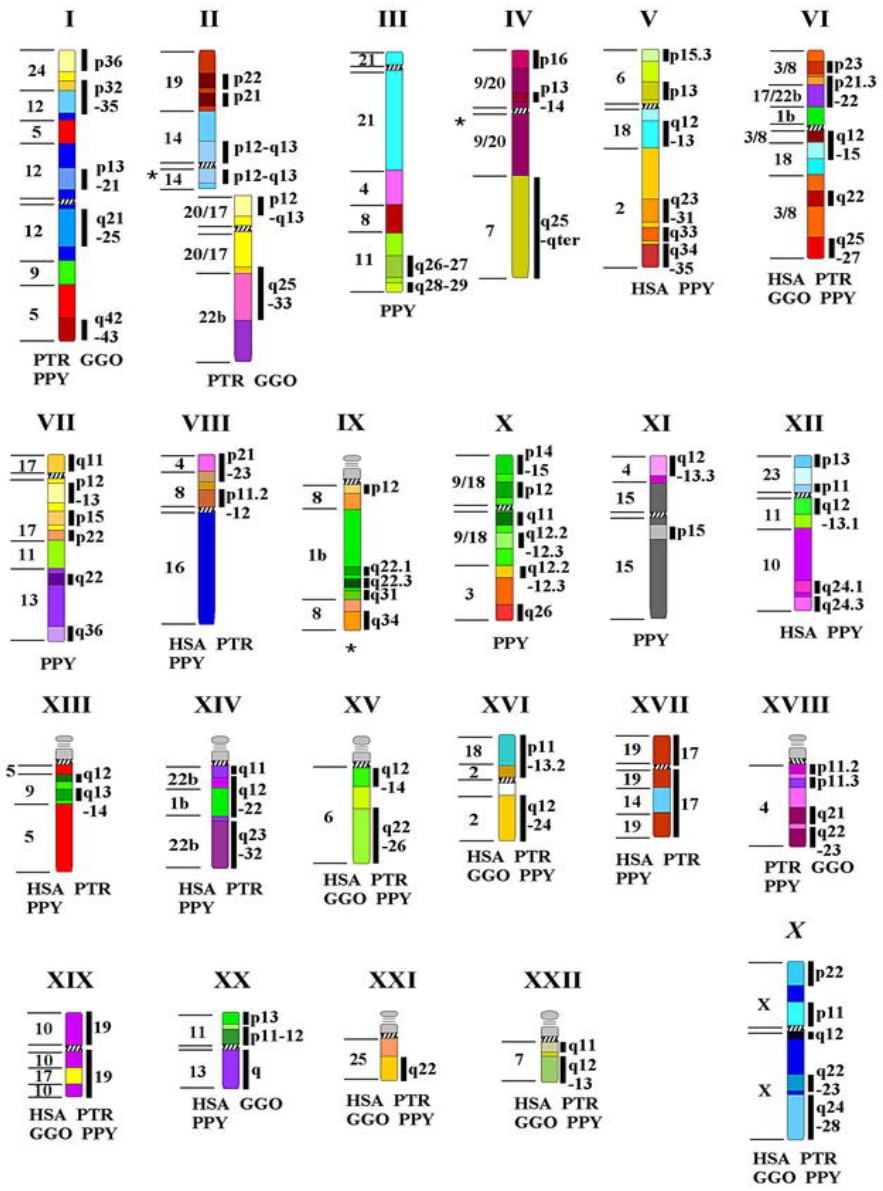



Figure 1.2. Reconstruction of the ancestral human and great ape karyotpye. Below each chromosome, except for IIp, IV and IX, the species showing the proposed ancestral form is indicated. At the left of each chromosome are the homologous gibbon chromosome regions. At the right are segments and bands that were defined by fragmented hybrids. (From Müller and Wienberg, 2001)

Table 1.3. Ancestral chromosomal derived forms bar code patterns (Müller and Wienberg, 2001).

| *Chromosome* | | *HSA* | *PTR* | *GGO* | *PPY* |
|---|---|---|---|---|---|
| I | | expansion[a] | ancestral | ancestral | ancestral |
| II | IIp | fusion | /[b] | / | / |
| | IIq | | ancestral | ancestral | 1 pe |
| III | | 2 pe | 2 pe | 2 pe | ancestral |
| IV | | / | / | / | / |
| V | | ancestral | 1 pe | 1 t | ancestral |
| VI | | ancestral | ancestral | ancestral | ancestral |
| VII | | 1 pe +1 pa | 1 pe +1 pa | 1 pe | ancestral |
| VIII | | ancestral | ancestral | 1 pe | ancestral |
| IX | | / | / | / | / |
| X | | 1 pa | 1 pa | 1 pe +1 pa | ancestral |
| XI | | 1 iv | 1 iv | 1 iv | ancestral |
| XII | | ancestral | 1 pe | 1 pe | ancestral |
| XIII | | ancestral | ancestral | 1 pa | ancestral |
| XIV | | ancestral | ancestral | 1 pe | ancestral |
| XV | | ancestral | ancestral | ancestral | ancestral |
| XVI | | ancestral | ancestral | ancestral | ancestral |
| XVII | | ancestral | ancestral | 1 t | ancestral |
| XVIII | | 1 pe | ancestral | ancestral | ancestral |
| XIX | | ancestral | ancestral | ancestral | ancestral |
| XX | | ancestral | 1 pa | ancestral | ancestral |
| XXI | | ancestral | ancestral | ancestral | ancestral |
| XXII | | ancestral | ancestral | ancestral | ancestral |
| ***X*** | | ancestral | ancestral | ancestral | ancestral |

Roman nomenclature was used for homologous chromosomes of great apes. HSA: *Homo sapiens*, human; PTR: *Pan troglodytes*, chimpanzee; GGO: *Gorilla gorilla*, gorilla; PPY: *Pongo pygmaeus*, orangutan. a: human chromosome 1 showed an expansion of pericentromeric heterochromatin, compared with other great apes. The bar code patterns are identical. b: No ancestral chromosome form could be postulated from bar code patterns. “pe” indicates pericentric inversion, “t” translocation, “pa” paracentric inversion, “iv”: inversion.

### 1.2.3. Molecular cytogenetic dissection of human chromosome 3 evolution

Chromosome painting with human chromosome-specific probes revealed an ancestral association of human chromosome 3 and 21 homologs. It is found in a wide array of species from a number of mammalian orders including insectivores, artiodacyls, carnivores, tree shrews, and lower primates (Table 1.1). Available gene mapping data suggests that both chromosomes are also associated in marsupials (Comparative Genome organization of vertebrates, 1996). The syntenic association of 3 and 21 may be ancestral for all placental mammals and is the largest conserved syntenic block in mammals. Chromosome painting results identified the fission of the 3/21 association as a genomic landmark for higher Old World primates and humans after the divergence of prosimians and New World monkeys (Richard *et al.,* 1996; Müller *et al.,* 1997).

Intrachromosomal rearrangements of the syntenic block 3/21 were studied by using probes derived from chromosomal subregions with a resolution of up to 10-15 Mb (Müller et al., 2000). Chromosome painting probes derived from tree shrews, which previously have been shown to be homologous to different regions of human chromosomes 3/21, or subregional probes from rearranged human/rodent somatic cell hybrids and YAC DNA probes were used for comparative mapping. The results demonstrate that the gene order may be highly rearranged although the synteny of human chromosome 3 has been highly conserved in various mammals. The ancestral form was proposed to be conserved in the Bornean orangutan subspecies with two subsequent inversions occurring in the phylogenetic line leading to African apes and humans. A pericentric inversion and centromere shift are needed to derive the ancestral chromosome of Old World monkeys (Figure 1.3).

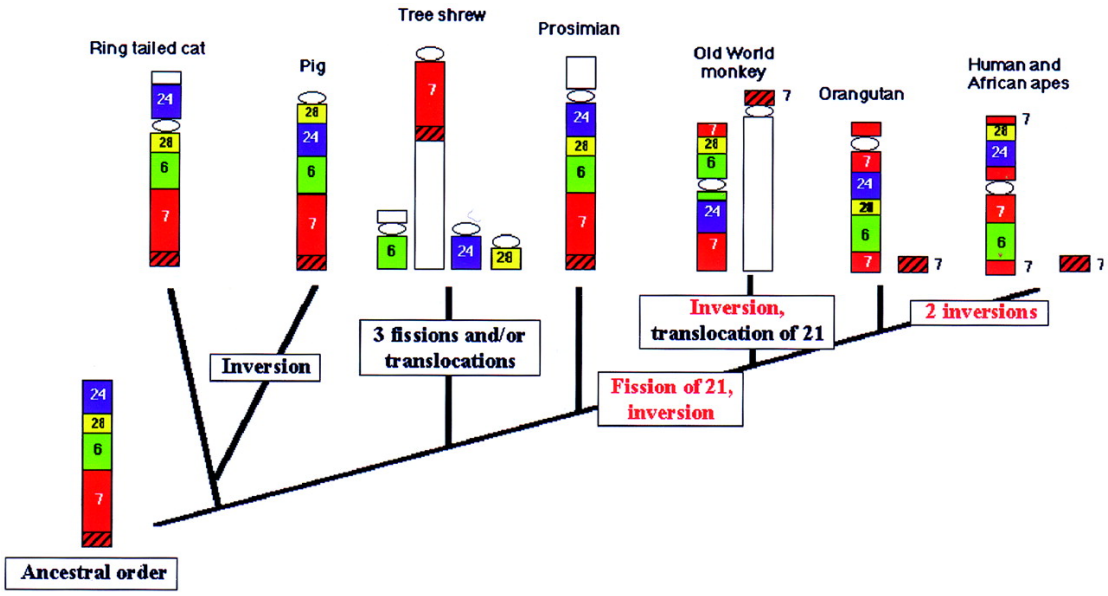



Figure 1.3. Summary and interpretation of the hybridization results of tree shrew chromosome 6-(green), 7-(red), 24-(blue) and 28-(yellow) on various mammalian species. The segment homologous to human 21 is hatched. The ancestral order of chromosome segments for carnivores, artiodactyls and primates is found in the ring-tailed cat and prosimians. The origin of Old World monkeys, apes and humans is marked by a fission of human chromosome 21 and 3 synteny and an inversion. The Bornean orangutan has conserved the ancestral order for all higher Old World primates. The Old World monkeys are derived by an inversion from the ancestral form. African apes and humans are phylogenetically linked by two inversions. (From Müller *et al.*, 2000)

Comparative mapping of more than 100 region-specific clones from human chromosome 3 in Bornean and Sumatran orangutans, siamang gibbons, Old and New World monkeys provided the possibility to reconstruct ancestral simian and hominoid chromosomes (Tsend-Ayush *et al.*, 2004). The human chromosome 3 consists of four large DNA segments (arrows marked in different color in figure 1.4) which have been conserved in their entirety in Bornean orangutan (*Pongo pygmaeus*, PPY) chromosome 2. The simian ancestral chromosome form is most similar to chromosome 1 of the Old World monkey *Presbytis cristata*. One shared and two independent inversions derive Bornean orangutan chromosome 2 and human chromosome 3 from the ancestral simian form, implying that neither Bornean orangutan nor humans have conserved the ancestral chromosome form. In the New World monkey *Callithrix geoffroyi* (CGE) and siamang *Hylobates syndactylus* (HSY), the ancestor diverged on multiple chromosomes. Centromere repositioning has occurred in the homologs of Old and New World monkeys.

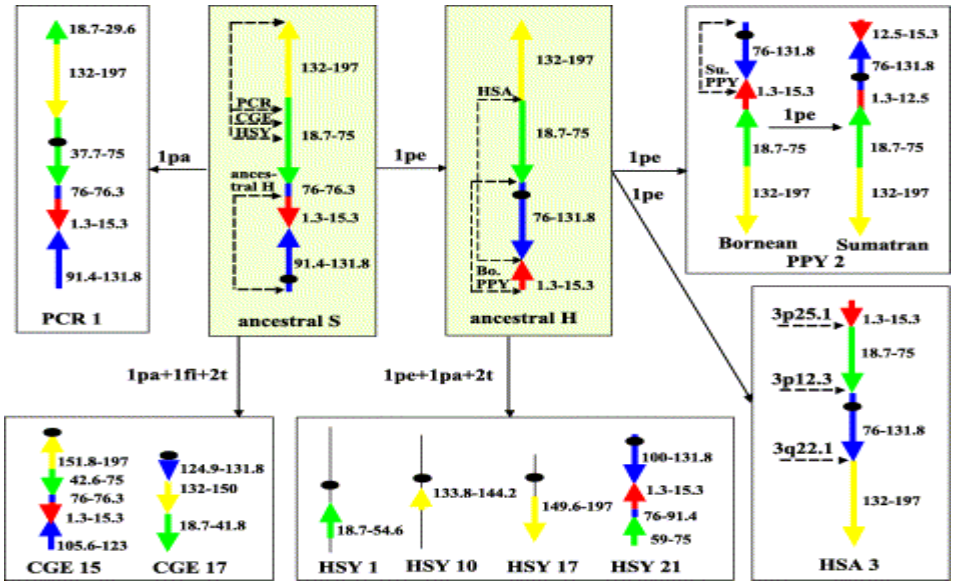



Figure 1.4. Evolutionary history of human (HSA) chromosome 3 in primates based on YAC hybridization results. Ancestral S and H represent hypothesized HSA3 ancestors for the simian primates and the hominoid family. “pe” indicates pericentric inversions, “pa” paracentric inversions, “fi” fissions and “t” translocations. The numbers at the right side of a given segment indicate the minimum size (Mb intervals) of the homologous region on HSA3, the direction (arrowhead) corresponds to the order of the hybridized YACs on the human chromosome. The thin black lines represent segments of a given primate chromosome, which are not homologous to HSA 3. The centromere position is indicated by a black dot on the chromosome. (From Tsend-Ayush *et al.*, 2004)

## 1.3. Segmental duplications and evolution of the primate genome

### 1.3.1. Features of segmental duplications

Segmental duplications are large (from 1 to >200 kb), nearly identical copies (90-100% sequence identity) of genomic DNA that are present in at least two locations in the human genome. They include intrachromosomal duplications which tend to occur in euchromatic regions and interchromosomal duplications which tend to be located in pericentromeric or subtelomeric regions. Segmental duplications contain both high-copy number repeats and gene sequences with intron-exon structures. Segmental duplications have been identified on every human chromosome with some chromosomes, i.e. the Y, showing peculiar enrichments for these types of duplication (Table 1.4). The identification and characterization of segmental duplications are largely based on computational and fluorescence *in situ* hybridization analysis of the human genome. Both methods indicate that approximately 5% of the human genome are composed of segmental duplications that have emerged during the past 35 million years of evolution (The international Human Genome Sequencing Consortium, 2001;

Cheung et al., 2001).

Table 1.4. Distribution and percentage of segmental duplications in the human genome. (From Samonte and Eichler, 2002)

| *Chromosome* | *Intrachromosomal (%)* | *Interchromosomal (%)* | *All (%)* |
|---|---|---|---|
| 1 | 1.4 | 0.5 | 1.9 |
| 2 | 0.1 | 0.6 | 0.7 |
| 3 | 0.3 | 1.1 | 1.1 |
| 4 | 0.0 | 1.0 | 1.0 |
| 5 | 0.6 | 0.3 | 0.9 |
| 6 | 0.8 | 0.4 | 1.1 |
| 7 | 3.4 | 1.3 | 4.1 |
| 8 | 0.3 | 0.1 | 0.3 |
| 9 | 0.8 | 2.9 | 3.7 |
| 10 | 2.1 | 0.8 | 2.9 |
| 11 | 1.2 | 2.1 | 2.3 |
| 12 | 1.5 | 0.3 | 1.8 |
| 13 | 0.0 | 0.5 | 0.5 |
| 14 | 0.6 | 0.4 | 1.0 |
| 15 | 3.0 | 6.9 | 6.9 |
| 16 | 4.5 | 2.0 | 5.8 |
| 17 | 1.6 | 0.3 | 1.8 |
| 18 | 0.0 | 0.7 | 0.7 |
| 19 | 3.6 | 0.3 | 3.8 |
| 20 | 0.2 | 0.3 | 0.5 |
| 21 | 1.4 | 1.6 | 3.0 |
| 22 | 6.1 | 2.6 | 7.5 |
| X | 1.8 | 3.2 | 5.0 |
| Y | 12.1 | 16.0 | 27.4 |
| Unknown | 0.0 | 0.5 | 0.5 |
| **Total** | 2.0 | 1.5 | 3.3 |

The calculation was based on the finished human genome sequence of September 2000 (http://genome.UCSC.edu/). The analysis excludes duplications with identities >99.5% to avoid artefactual duplicates caused by incomplete assembly of working draft sequence. There is some overlap between the interchromosomal and intrachromosomal sets. By these criteria, only 3.3% of the human genome is duplicated. However, estimates based on FISH and computer analysis indicate that the final amount will be ~5%.

Phylogenetic and comparative analyses of several regions strongly support a two-step model of duplication (Figure 1.5). Acceptor regions of the genome acquire segments of 1-200 kb from different genomic regions (donor loci) through a process of duplicative transposition. A series of such events create a mosaic of duplicated segments that originate from different regions of the human genome. Secondary events duplicate portions of this mosaic structure to other intrachromosomal or interchromosomal sites. Subsequently rearrangements (deletions and inversions) alter the structure of these regions.

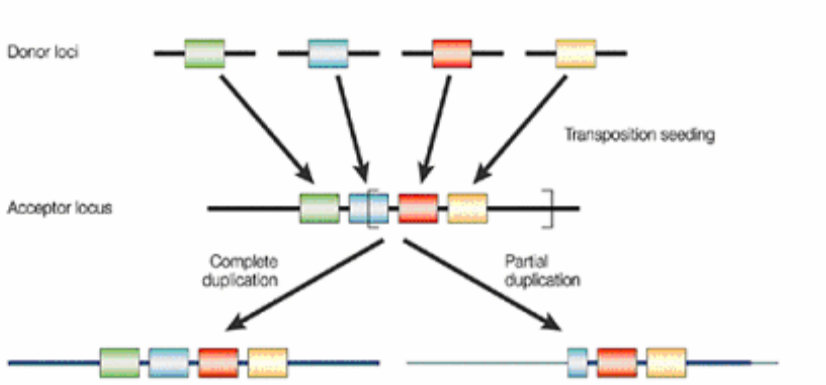



Figure 1.5. Model of segmental duplication. (From Samonte and Eichler, 2002)

## 1.3.2. Consequences of segmental duplications

There are several potential consequences of recent segmental duplications. First, when the duplicated genomic segment is non-functional, this only leads to the accumulation of unprocessed pseudogenes (Lynch and Conery, 2000). Many of the duplicated segments appear to contain only partial gene sequences and miss the regulatory elements which are necessary for expression. However, careful comparison of expressed sequence tags (ESTs) and genomic sequence databases showed that although the transcripts frequently contain premature stop codons, many duplicated segments are transcribed, sometimes in a tissue-specific fashion (Lyer *et al.*, 1996). The function of such transcripts is not well known although they are assumed to be non-functional. Secondly, there are several examples in which the transcripts are chimeric; transcription proceeds across adjacent duplications that originated from different regions of the genome (Eichler, 2001). The production of such mosaic transcripts which consist of exons from different genes might contribute to the increased complexity of the human proteome (Venter *et al.*, 2001). Third, segmental duplications provide the possibility of evolution of new functions. It has been argued that duplicating a copy of a gene or genomic segment would release the gene copy

from the selective constraints and provide the possibility to acquire a modified or improved function (Samonte and Eichler, 2002). A novel hominoid gene family, the morpheus gene family, was discovered in half of the human duplications of LCR16a, which is a 20 kb segment. 15-30 copies are dispersed throughout 15 Mb of human chromosome 16 and the chimpanzee homolog. The average coding sequence divergence (15-20%) among different copies of the morpheus gene family between human and chimpanzee far exceeds the amount of intronic sequence divergence (1-2%) and most of coding sequence changes (>95%) resulted in amino-acid changes. Analysis of this gene family among hominoids shows that the main episode of enhanced amino-acid replacement occurred after the separation of the human, chimpanzee and gorilla lineages from the orangutan and the evolution of this duplication is under positive selection. It is possible that the rapid evolution of the morpheus gene family and genes derived from other segmental duplications have been crucial for the adaptation of humans to different environments (Johnson *et al.*, 2001; reviewed by Samonte and Eichler, 2002).

### 1.3.3. Genome plasticity and chromosomal rearrangements

Segmental duplications have occured at many different times during primate evolution. Differences in copy number and location of duplicated segments have been observed for many chromosomes, particularly between the genomes of humans and great apes (Jackson *et al.*, 1999; Horvath *et al.*, 2000). Structural polymorphisms of duplicated segments that range from 30 kb to 1 Mb have been identified in the pericentromeric regions of human chromosomal regions 16p11 and 15q11. The most proximal sequenced portion of human chromosome 22 has been shown to be the result of a human-specific duplication event that involved the transposition of this segment from human chromosome 14. This occurred after the separation of chimpanzee and humans (Bailey *et al.*, 2002b). Two copies of the Charcot-Marie-Tooth neuropathy type 1A repeat sequence (*CMT1A-REP*) have been identified in both chimpanzees and humans, but only the single ancestral locus is present in gorillas, which indicates that the locus has been duplicated in a common ancestor of chimpanzees and humans (Keller *et al.*, 1999). Some duplicated loci, for example, the segment including the creatine transporter gene *SLC6A8* and the adrenoleukodystrophy gene *ABCD1* at Xq28, seem to be restricted to humans and the African apes, but there is only a single copy in the orangutan and Old World monkeys (Eichler *et al.*, 1996).

Several studies have shown that highly homologous sequences at distant loci predispose to illegitimate recombination, which can lead to large-scale chromosome

rearrangements such as deletions, duplications, paracentric inversions and translocations (reviewed by Stankiewicz and Lupski, 2002). Most large-scale chromosomal rearrangements between humans and great apes are pericentric inversions. In addition, there are a reciprocal translocation involving the gorilla homologs of human chromosome 5 and 17 and a fusion of two acrocentric ancestral chromosomes to human chromosome 2. It has been postulated that these large-scale rearrangements create genetic barriers that lead to stasipatric speciation. FISH analyses combined with sequence and computational analyses have been widely used for the characterization of evolutionary chromosomal rearrangements. Comparative mapping of human chromosome 7 and the orthologous mouse regions showed the presence of large low-copy repeats at the inversion breakpoints between humans and mouse chromosomes (Valero *et al.,* 2000). Intriguingly, a ~250 kb sequence has been identified to be duplicated near both breakpoints of the gorilla specific translocation [corresponding to human t(5;17)] (Stankiewicz *et al.,* 2001). Although the cause-and-effect relationship between segmental duplications and evolutionary rearrangements remains to be elucidated, two models have been proposed. Duplication-driven chromosomal rearrangements model proposes that segmental duplications provide templates for non-allelic homologous recombination events and, thus, facilitate large chromosomal rearrangements that trigger speciation. Alternatively, the segmental duplications might be products of the rearrangement process itself. Examination of the structural and functional features of evolutionary breakpoints will provide a better understanding of the association between segmental duplications and genome evolution.

## 1.4. Aims of this research

A previous study revealed three breakpoints between human chromosome 3 and Bornean orangutan chromosome 2 (Tsend-Ayush *et al.,* 2004). The breakpoint at human chromosome 3p25.1 was located in YAC clone 772d2. The breakpoint at 3p12.3 was located in YAC clone 905d2. The breakpoint at 3q21.3 was located in the 2 Mb interval between YAC clones 896g9 and 914h1. This interval overlaps with two other independent evolutionary breakpoints, namely the translocation breakpoint between human chromosome 3 and the siamang gibbon genome as well as the inversion breakpoint between human chromosome 3 and *P. cristata* chromosome 1. In order to further study these evolutionary breakpoints and to narrow down the breakpoint intervals, human chromosome 3 region specific BAC clones will be selected by database analyses and hybridized to primates metaphase chromosomes.

Higher resolution PCR mapping will be performed on flow-sorted primate chromosomal libraries using primers designed from human genomic sequence. *In silico* sequence analyses will be used to characterize the gene contents and the sequence features such as repetitive elements and segmental duplications at evolutionary breakpoints. Characterization of new genes or gene isoforms located at breakpoint regions provide candidate genes for phenotypic differences between primate species. Pathological rearrangements are considered to be a temporary snapshot of ongoing genome evolution. Therefore, the mechanism underlying evolutionary changes may also be responsible for human pathology. In this light, the detailed analysis of evolutionary breakpoints can provide a better understanding of the evolutionary processes remodeling complex genomes, such as the forces driving chromosomal rearrangements and structural features promoting breakage, as well as new insights in human pathology.

## 2.1. Common materials

All chemicals were purchased from Roth if not stated otherwise. All solutions were prepared with Millipore-filted $H_2O$ if not described otherwise. Common materials for bacterial cultures, gel electrophoresis, sequencing as well as DNA and RNA sources used in this research are listed in table 2.1. The separation of DNA weight markers on agarose gels is shown in figure 2.1.

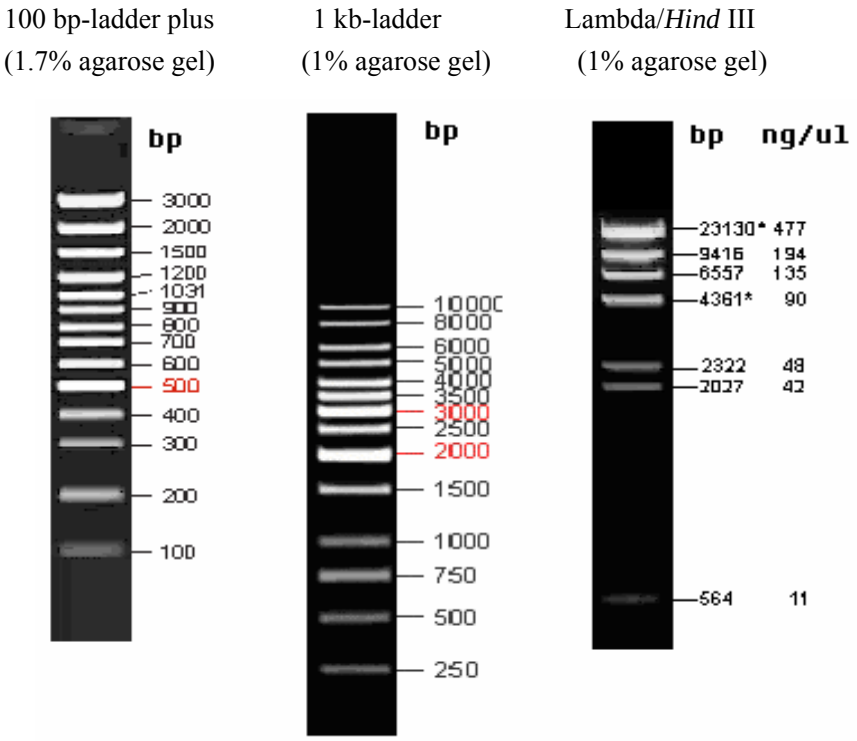



Figure 2.1. Separation of DNA weight molecular markers on agarose gels (adapted from www.fermentas.com)

## 2.2. Materials for DNA and RNA isolation

Kits and reagents used for DNA and RNA isolation are listed in table 2.2.

## 2.3. Materials for PCR (Polymerase Chain Reaction) analysis

### 2.3.1. Standard PCR

The reagents needed for PCR are listed in table 2.3. Primer pairs were obtained from MWG Biotech and are listed in tables 2.4 to 2.7.

Table 2.1. Common materials used in this project.

| ***Materials*** | | ***Description*** |
|---|---|---|
| Bacterial cultures | LB (Luria –Bertani) Medium | 10 g NaCl, 10 g Bacto Tryptone, 5g Yeast Extract, add 1L $H_2O$, autoclaved |
| | LB-Agar for plates | Add 15 g/L Agar to LB-Medium, autoclaved |
| | S.O.C- medium | 2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose (provided with TOPO TA cloning kit) |
| | X-Gal | 40 mg/ml in dimethylformamide |
| | Chloramphenicol | solution in 100% ethanol, sterilized with sterile filter, aliquots at -20°C stock concentration: 34mg/ml ; work concentration: 17 µg/ml |
| | Kanamycin | solution in $H_2O$ bidest, sterilized with sterile filter, aliquots at -20°C stock concentration: 20 mg/ml; work concentration: 50 µg/ml |
| Gel electrophoresis | Agarose | Seakem ® LE-Agarose (Cambrex; Cat.No. 50004) |
| | 1x TAE buffer | 4 mM Tris-acetate pH 7.8, 1mM EDTA |
| | 10x loading buffer (orange G) | 20 ml glycerol (87%), 1ml 1M Tris, pH 7.5, 0.1g Orange G (Sigma; Cat. No. O-7252); add $H_2O$ up to 100 ml |
| | Ethidium bromide | stock solution: 10 mg/ml |
| | Molecular weight markers | Gene Ruler™ 100bp DNA ladder plus; Gene Ruler™ 1 kb DNA ladder; Lambda DNA/Hind III Marker; (MBI Fermentas SM0321; SM0311; SM0101) |
| Sequencing | CEQ Dye Terminator Cycle Sequencing (DTCS) Quick Start Kit (Beckman Coulter; Cat.No. P/N 608120), CEQ sample plate (Beckman Coulter; Cat.No. P/N 609801) | |
| | 3 M Sodium acetate pH5.2 (Sigma; Cat. No. S 7899) | |
| | 100 mM $Na_2EDTA$ pH8.0, prepared from 0.5M $Na_2EDTA$ (Sigma; Cat.No. E 7889) | |
| DNA & RNA sources | Primates chromosome specific libraries: DOP PCR products of flow sorted chromosomes (provided by Prof. Malcolm Ferguson-Smith and Dr. Fengtang Yang) | |
| | Human specific RNA (1µg/µl) from following tissues [Human total RNA Master Panel II (Clontech; Cat.No.K4008-1)]: bone marrow, brain (cerebellum), fetal brain, fetal liver, heart, kidney, lung, placenta, prostate, salivary gland, skeletal muscle, spleen, testis, thyroid gland, trachea, uterus, colon with mucosal lining, small intestine, spinal cord and stomach. | |

Table 2.2. Kits and reagents used for DNA and RNA extraction.

| *Application* | | *Kits or Reagents* | *Company* | *Cat. No.* |
|---|---|---|---|---|
| DNA | Gel extraction of PCR fragments | NucleoSpin® Extract | Macherey-Nagel | 740590.250 |
| | Plasmid DNA isolation | NucleoSpin® plasmid | Macherey-Nagel | 740588.250 |
| | BAC DNA isolation | PhasePrep™ BAC DNA Kit | Sigma | NA 0100 |
| RNA | RNA isolation from tissue | TRIzol® Reagent | Invitrogen | 15596-026 |
| | RNA from cell pellets | NucleoSpin® RNA II | Macherey-Nagel | 740955.250 |

Tabel 2.3. Reagents for standard PCR.

| *Reagent* | *Description* | *Company* | *Cat. No.* |
|---|---|---|---|
| *Taq* DNA Polymerase | Isolated from *Thermus aquaticus* YT1, heat-stable; 5 U/µl | Invitrogen | 18038-042 |
| 10x PCR buffer minus Mg | 200 mM Tris-HCl (pH8.4); 500 mM KCl | Invitrogen | Provided with Taq DNA polymerase |
| $MgCl_2$ | 50 mM | Invitrogen | Provided with Taq DNA polymerase |
| dNTPs | 10 mM of each dATP, dCTP, dGTP, dTTP; diluted from 100 mM stocks | MBI Fermentas | R0182 |

Table 2.4. Primer pairs used for PCR mapping of BAC RP11-803B1 fragments in the human and orangutan genomes. The fragment "a" is most proximal one and the fragment "o" is most distal.

| *PCR fragment* | *Forward primer sequence (5'-3')* | *Reverse primer sequence (5'-3')* | *Position (bp) in BAC RP11-803B1* |
|---|---|---|---|
| 803B1a | tgcatggctaattccatcaa | tgaacagaacttcctgagaa | 1721-1880 |
| 803B1b | ctctgtgaggagtgcataac | tgtccacatcctgatcctag | 19781-19960 |
| 803B1c | atacatgagcgagagcagtt | gtagctccgtgtaccaatca | 42061-42220 |
| 803B1d | tcagtcttcttcctgtgact | gcttctaagctctgtgtgga | 78031-78240 |
| 803B1e | ctgagagacctgtccgaaat | acaggtttgctgtacacttg | 97561-97760 |
| 803B1f | tagcgaccctgagtgtttct | ccaggatcacagtgcatttg | 98101-98260 |
| 803B1h | atcaacagagcatctccaag | ctccgggtcaaattagccat | 106841-107030 |
| 803B1i | acttggaatgtgctaggaca | tcaagacctgtatgtgcagt | 111531-111720 |
| 803B1j | tgaggaggctgaagagatct | ctgaaagttccagccttctt | 112071-112230 |
| 803B1k | ctgtacaagcgatgatgtag | agatgttagctgtcttgctg | 112341-112530 |
| 803B1m | acaggaagagtgagtgactc | atgctggcagtaagccataa | 124081-124320 |
| 803B1n | agtcagagctacgtgattcc | gggttgtccagaagaaagca | 131671-131860 |
| 803B1o | atacagacccacagagtcta | gatgatgagcaggttcctga | 139741-139880 |

Table 2.5. PCR fragments of BAC RP11-642N14 for FISH mapping. Fragment "A" is most proximal and fragment "D" is most distal.

| *Fragment* | *Forward primer sequence (5'-3')* | *Reverse primer sequence (5'-3')* | *Position(bp) in BAC RP11-642N14* |
|---|---|---|---|
| 642N14A | atttcctctgtgacccttct | gcacctccaatctctttcaa | 9-10120 |
| 642N14B | attgtgggtttgaagggatc | aggcatgctgccttttctat | 20190-30200 |
| 642N14C | ttagctcaccccaattcaga | ttaggccacttgctacagaa | 39691-50010 |
| 642N14D | ccacatgagctctattcatc | ggaaaattctgcctttccag | 120551-130850 |

Table 2.6. Primer pairs used for PCR amplification of BAC RP11-93K22, RP11-77P16, and RP11-687B8 fragments from human, orangutan, and gibbon chromosomal libraries.

| *PCR* | *Forward primer (5'-3')* | *Reverse primer (5'-3')* | *Position in BAC* |
|---|---|---|---|
| 93K22d | tgt att ggc ttc aca gca ca | agc act tta cag tta gca cc | 162321-162520 |
| 93K22s | tcg gac acc tca tca tcc at | caa tat tca ggg tgc agg ct | 143701-143920 |
| 93K22c | tct gac tcc act tag acc ac | cta cat gga gaa ggt gat ca | 140561-140680 |
| 93K22e | ctg cac cac ctc tgg aga t | gtg tgg atc tcg gtc ttt gt | 134002-134140 |
| 93K22g | gcc tgc tct ctg tga aat at | gct tgg tca gct agc ctt ct | 132251-132390 |
| 93K22h | ctc aca tgc cta aag gaa ga | ctg aag tcg agt ttg tgg tc | 131099-131270 |
| 93K22r | atg aaa gac cag tgg act tg | ttc tca ctc tgc act gtg tc | 123681-123870 |
| 93K22v | gct ggt gca gtc ata gga aa | tac agc aga gag aag cag ca | 115451-115670 |
| 93K22q | gga cta tct gag act gga tt | atg ttg cca cag ttc aca gt | 112511-112790 |
| 93K22p | agt gac ctg cag gtt gga ta | aga cag cag gct atg cta tg | 109611-109820 |
| 93K22l | ttc ttc act cac agg cca tg | atc tga atg ccc atg ctg ct | 108171-108340 |
| 93K22k | cct cca ctg tac agt tga ca | act tga ggc aca tta tgc tg | 98321-98450 |
| 93K22j | aca tgc agt cgc ata gca ga | ttc acg atg act ggg tag tg | 86231-86380 |
| 93K22i | aac cga tgt cag gca agg ac | ctg ttc ctg tcc atg tgc ct | 73101-73230 |
| 93K22b | tag gaa cag cag agg tct ct | tgc cag aga gtc gct aac aa | 60191-60339 |
| 93K22a | atg ctg caa tat cct gtg cc | cct atg agc ttc aag cca ac | 991-1160 |
| 77P16a | tgc agt gct gac tca atg gt | ctc atc aga ttc tga gct gc | 2591-2760 |
| 77P16b | act gag tac tgg gtc atc ag | tgt cat tgg aga gga gcg at | 49991-50094 |
| 77P16d | ctt acc tgg gat ggc agt ga | act cac aac tgc tgc tga tc | 61351-61520 |
| 77P16f | atc tcc agg tga gga ttc ag | gag cac agt cgt agc ctt aa | 79381-79500 |
| 77P16g | gaa agc aag ctg gga gtc aa | atg aga aat gcg gaa tgc ca | 90401-90560 |
| 77P16p | gga gat cat gct gct gtg at | atg cac gat att gcg tgt gg | 93871-94096 |
| 77P16h | gtc aca tga gga ctt gac at | ctg atg cca agg ttg gct aa | 104981-105101 |
| 77P16j | ttg aca gag aca gtt gct tg | aag act cca tag cta tgt tg | 106321-106540 |
| 77P16k | agg aga gaa ggt agg taa tc | ata gtc act cat tag cat gg | 107781-107960 |
| 77P16i | tct tcc tag gct cac tgt ct | tga cat gcc aag gaa gct gt | 112231-112420 |
| 687B8d | cgt cag gta aga tgt tct gt | ttc cca cag gct tgc tct ct | 2751-2870 |
| 687B8e | tgc cta gag agc tag ctt ct | cgt tca agg gca tta agg tc | 11821-11990 |
| 687B8g | gca gcc aga tca gca aac tt | gct aac ttg ctc tcc tgt tc | 16741-16910 |

Table 2.7. Primers used for human and mouse *ROBO2* analyses.

| ***Number*** | ***Sequence (5'- 3')*** | ***Tm (°C)*** |
|---|---|---|
| 1 | tgt agc cac aca agt ata agt | 58 |
| 2 | ctc cca gac aga gag tgg | 58 |
| 3 | gag tct gct gat gtt tac ac | 58 |
| 4 | agc tct gta gct gac agg c | 60 |
| 5 | gat gtg gta gtc gca gct g | 60 |
| 6 | gca acg atc tga tct ctt gg | 60 |
| 7 | act tat act tgt gtg gct aca | 60 |
| 8 | ctc cag tac aac cag atg ct | 60 |
| 9 | agc aat agc tgg cag aca g | 58 |
| 10 | act aca gag cga atg gct g | 58 |
| 11 | agc atc tgg ttg tac tgg ag | 60 |
| 12 | agt ggg cag tgg act gct gt | 62 |
| 13 | aca gca gtc cac tgc cca ct | 60 |
| 14 | gga act cca tgt tat agt ctg | 60 |
| 15 | tac agc aag ccc agc ttc c | 60 |
| 16 | tac acc agc att atc caa cc | 58 |
| 17 | ccg cgg ggg aaa gtc ctc ctg gcg aa | 67.8 |
| 18 | aat gca atg gcc aga aga ca | 58 |

## 2.3.2. Long range PCR

The "Expand Long Template PCR System" from Roche Applied Science (Cat. No. 1 681 834) was used to amplify fragments of ~10 kb. This system is composed of a unique enzyme mix containing thermostable *Taq* DNA polymerase and a proofreading polymerase. It amplifies fragments up to 40 kb from λ DNA. Buffers are provided with the kit.

### 2.3.3. RT PCR

Reverse transcription PCR was performed with SuperScript™ III Reverse Transcriptase to synthesize first-strand cDNA. The subsequent PCR to amplify larger amounts of cDNA was performed with the standard PCR protocol. The materials needed for reverse transcription are listed in table 2.8.

Table 2.8. Reagents needed for reverse transcription.

| ***Reagent*** | ***Company*** | ***Cat. No.*** |
|---|---|---|
| SuperScript™ III Reverse Transcriptase (200 U/μl) | Invitrogen | 18080-093 |
| 5x First-Strand Buffer | Invitrogen | provided with reverse transcriptase |
| 0.1 M DTT | Invitrogen | provided with reverse transcriptase |
| 10 mM dNTP mix | Invitrogen | 18427-013 |
| RNAseOUT™ Recombinant Ribonuclease Inhibitor (40 U/μl) | Invitrogen | 10777-019 |
| RNase H (5 U/μl) | New England Biolabs | M0297S |

## 2.4. Materials for 5'-RACE (rapid amplification of cDNA ends)

### 2.4.1. First-strand cDNA synthesis

The BD SMART™ RACE cDNA amplification kit from BD Biosciences Clontech (Cat. No.K1811-1) was applied to synthesize 5'-RACE-Ready cDNA samples from total RNA.

### 2.4.2. RACE

The BD Advantage 2 PCR Kit (Cat.No.639207) was used to perform RACE PCR reactions for specific genes from 5'-RACE-Ready cDNA samples.

## 2.5. Materials for molecular cloning

### 2.5.1. Cloning of PCR products shorter than 3 kb

The TOPO® TA Cloning Kit from Invitrogen (Cat. No. K 4500-01) was used to clone

short (< 3 kb) PCR products. Figure 2.2. shows the vectors and the sequences surrounding the cloning sites. Other materials needed for the cloning reaction and bacterial cultures are documented in the corresponding part of the "methods" chapter 3.

## 2.5.2. Cloning of long PCR products

The TOPO® XL PCR Kit from Invitrogen (Cat. No. K 4700-20) was used to clone long (3-10 kb) PCR products. The vector is shown in figure 2.2.

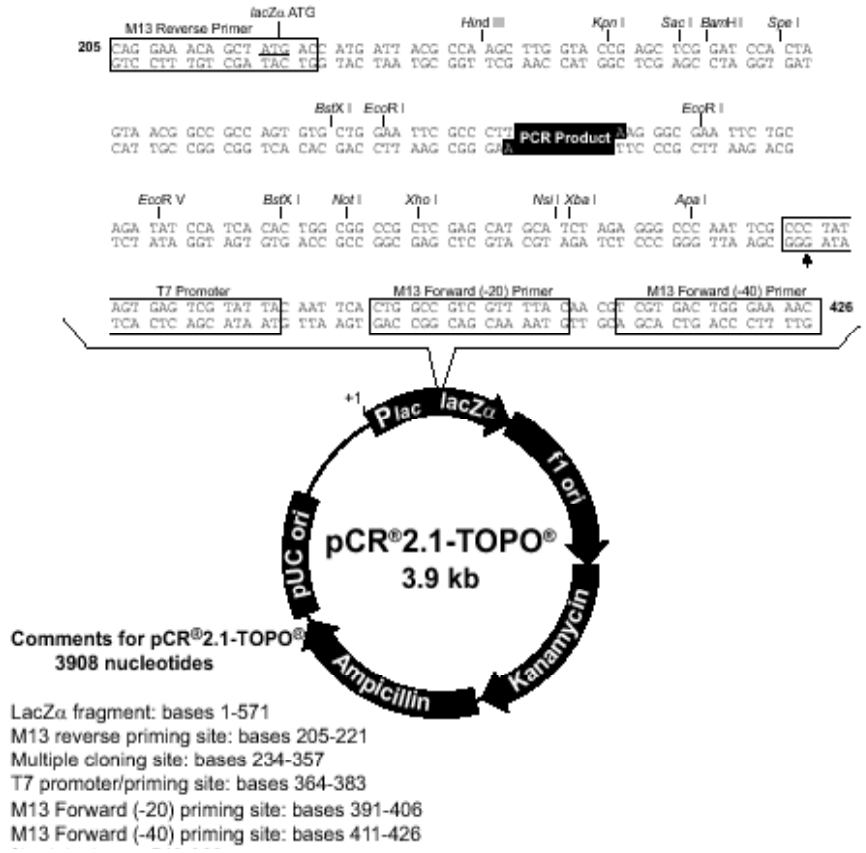
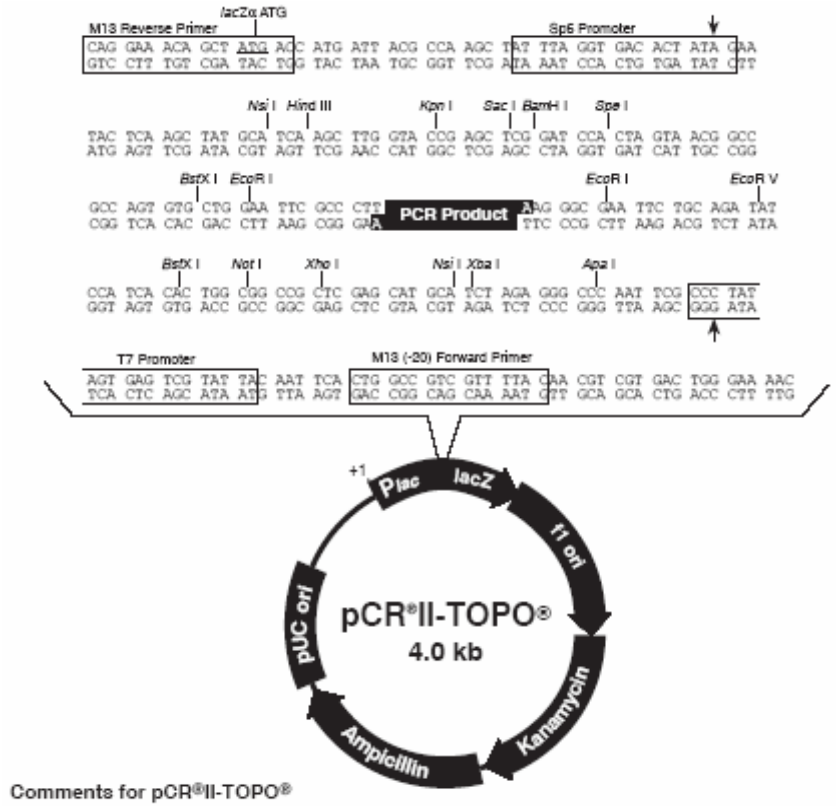
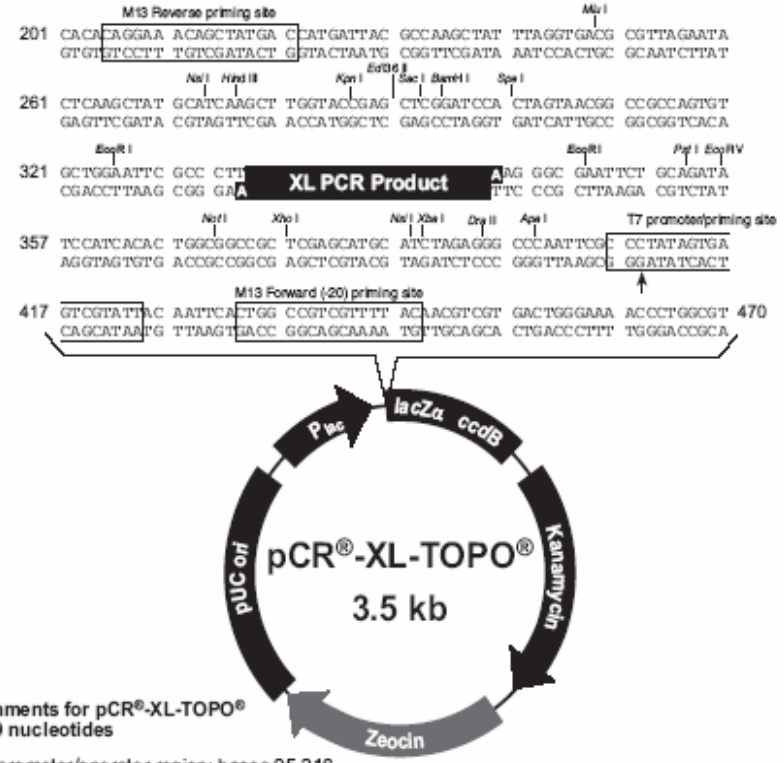



Figure 2.2. (Continued on next page) Vectors pCR®2.1-TOPO® and pCR®II-TOPO® for cloning short PCR products and vector pCR®-XL-TOPO® for cloning long PCR products. (From Invitrogen instruction manual).




**Comments for pCR®II-TOPO®**
**3973 nucleotides**

*LacZα* gene: bases 1-589
M13 Reverse priming site: bases 205-221
Sp6 promoter: bases 239-256
Multiple Cloning Site: bases 269-383
T7 promoter: bases 406-425
M13 (-20) Forward priming site: bases 433-448
f1 origin: bases 590-1027
Kanamycin resistance ORF: bases 1361-2155
Ampicillin resistance ORF: bases 2173-3033
pUC origin: bases 3178-3851




**Comments for pCR®-XL-TOPO®**
**3519 nucleotides**

Lac promoter/operator region: bases 95-216
M13 Reverse priming site: bases 205-221
Lac Za ORF: bases 217-576
Multiple Cloning Site: bases 248-399
TOPO® Cloning site: bases 336-337
T7 promoter priming site: bases 406-425
M13 Forward (-20) priming site: bases 433-448
Fusion joint: bases 577-585
*ccd*B lethal gene ORF: bases 586-888
Kanamycin resistance ORF: bases 1237-2031
Zeocin resistance ORF: bases 2238-2612
pUC origin: bases 2680-3393

## 2.6. Materials for FISH (fluorescence *in situ* hybridization)

### 2.6.1. Special reagents and solutions

Special reagents and solutions for FISH are listed in tables 2.9 and 2.10, respectively.

### 2.6.2. Probe sources for FISH analysis

1. Human chromosome 3 clones of the "RPCI-11 human BAC library" were obtained from the Resource Center Primary Database of the German Human Genome Project and the Children's Hospital Oakland-BACPAC Resources. Male blood was obtained via a double-blind selection protocol to create the RPCI-11 (Roswell Park Cancer Institute) library. Male blood DNA was partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector between EcoRI sites. For Segment 5, the same donor DNA was partially digested at the BamHI sites. The RPCI-11 library has 32.2x genomic redundancy.

   BAC clones of the "CITB Human D (CTD) BAC library" were obtained from ResGen (Invitrogen). The CITB (Caltech Institute of Technology BAC) Human D BAC library is composed of 5 segments. D1 uses the HindIII cloning site, while D2 through D5 use the EcoRI cloning site. The pBeloBAC11 vector was used for the construction of the Human D clones. The DNA source for all segments is human sperm (HSP). The average insert size for D1 is 129 kb with the average for segments D2-D5 being 173 kb. The CTD library represents a 17x coverage of the human genome.

   All BAC clones used in this research are listed in table 2.11. DNA extraction kits are listed in table 2.2.
2 Long-range PCR products were amplified from human BAC clones by expand long template PCR system (see 2.3.2). The products were separated on a 0.8% agarose gel and extracted by NecleoSpin® Extract kit. The primer pairs used for producing FISH probes are listed in table 2.5.

### 2.6.3. Metaphase chromosomes for FISH analysis

Metaphase chromosomes were prepared from EBV-transformed lymphoblastoid cell lines of humans, chimpanzee, Bornean x Sumatran orangutan hybrids, siamang gibbon, and silvered-leaf monkey.

Table 2.9. Special reagents for FISH.

| ***Reagent*** | ***Description*** | ***Company*** | ***Cat. No.*** |
|---|---|---|---|
| Biotin-16-dUTP | Tetralithium salt, 1 mM solution | Roche Applied Science | 1 093 070 |
| Digoxigenin-11-dUTP, alkali- stable | Tetralithium salt, 1 mM solution | Roche Applied Science | 1 570 013 |
| DNase I | RNase–free | Roche Applied Science | 776785 |
| DNA polymerase I (Kornberg polymerase) | Endonuclease–free | Roche Applied Science | 642 720 |
| dATP | Li-salt, 100 mM solution | Roche Applied Science | 1 051 140 |
| dCTP | Li-salt, 100 mM solution | Roche Applied Science | 1 051 458 |
| dGTP | Li-salt, 100 mM solution | Roche Applied Science | 1 051 466 |
| dTTP | Li-salt, 100 mM solution | Roche Applied Science | 1 051 482 |
| Human Cot-1 DNA | 1 mg/ml | Invitrogen | 15279-011 |
| Herring Sperm DNA | 10 mg/ml | Sigma | D 7290 |
| Formamid | deionized, aliquots at -20°C | Roth | P040.1 |
| Fluorescein avidin DCS (cell sorter grade) | 2 mg/ml in 10 mM HEPES, 0.5 M NaCl, pH8.0, 0.08% sodium azide | Vector | A-2011 |
| Biotinylated anti-avidin D | 0.5 mg reconstituted in 1 ml $H_2O$. aliquots at -20 °C | Vector | BA-0300 |
| Monoclonal anti-digoxin antibody DI-22 | mouse ascites fluid | Sigma | D 8156 |
| Cy™3-conjugated Affinipure Sheep anti-Mouse IgG (H+L) | 1.4 mg/ml | Dianova | 515-165-062 |
| DAPI | 1 mg/ml in $H_2O$, aliquots at -20°C | Sigma | D 9542 |
| Vectorshield® | Mounting media | Vector Laboratories | H-1000 |

Table 2.10. Special solutions used in FISH.

| ***Solution*** | ***Components*** |
|---|---|
| 10x nick translation buffer | 0.5 M Tris-HCl, pH7.5; 50 mM $MgCl_2$ ; 0.5 mg/ml BSA |
| dNTP-Mix | 0.5 mM dATP; 0.5 mM dCTP; 0.5 mM dGTP; 0.1 mM dTTP |
| Stop-Mix | 0.1% Bromphenolblue; 0.5% Dextran-Blue; 0.1M NaCl; 20 mM EDTA; 20 mM Tris-HCl, pH 7.5 |
| Sephadex G–50 | 25 g Sephadex–50 (Sigma-Aldrich; Cat.No. S-5897-100g) eliquilibrated in TE-buffer (10 mM Tris-HCl, pH8.0; 1 mM EDTA) containing 0.1% SDS |
| RNase A – stock | 20 mg/ml: dissolved in 10 mM Tris-HCl pH 7.5, 15 mM NaCl (DNase free) |
| Pepsin – Stock | 100 mg/ml (10%) in $H_2O$ |
| 20x SSC | 3 M NaCl/0.3 M Na-citrat (pH 7,0) |
| 10x PBS | 1.37 M NaCl; 26.8 M KCl; 43 mM $NaH_2PO_4$; 14,7 mM $KH_2PO_4$ |
| PBS/$MgCl_2$ | 95 ml 1x PBS + 5 ml 1M $MgCl_2$ |
| 1% formaldehyde | 1.3 ml formaldehyde (37%) in 50 ml PBS/$MgCl_2$ |
| FA/SSC | 70% formamid in 2x SSC |
| 2x hybridization mix | 4x SSC/20% dextransulfat (Sigma; Cat.No. D-8906) |
| 4x SSC/0,1% Tween | 200 ml 20x SSC, 2 ml Tween 20, add $H_2O$ to 1L |
| Blocking solution | 3% BSA in 4x SSC/0,1% Tween |
| Antibody dilution solution | 1% BSA in 4x SSC/0,1% Tween |
| Monoclonal anti-digoxin antibody DI-22 Stock | 1:10 diluted in PBS (10 mM phosphate, 150 mM NaCl, pH7.4) containing 0.1% sodium azide |
| DAPI/Vectorshield® | 1 µl DAPI (1 mg/ml) in 10 ml Vectorshield mounting medium |

Table 2.11. Human BAC clones from chromosome 3 used in this research (from 3p telomere to 3q telomere).

| *Name* | *Cytogenetic band* | *Accession Number* | *Name* | *Cytogenetic band* | *Accession Number* |
|---|---|---|---|---|---|
| RP11-72K22 | 3p25.1 | AC026102 | RP11-73I16 | 3p12.3 | AC133793 |
| RP11-135D4 | 3p25.1 | AQ381389 | RP11-58L10 | 3p12.3 | AC126467 |
| RP11-421B21 | 3p25.1 | AQ551922 | RP11-214P4 | 3p12.3 | AC024256 |
| RP11-316A10 | 3p25.1 | AC090937 | RP11-484E7 | 3p12.3 | AC067717 |
| RP11-616M10 | 3p25.1 | AC090954 | RP11-550F7 | 3p12.3 | AC026877 |
| RP11-321I9 | 3p25.1 | AC087590 | RP11-845C18 | 3p12.3 | AC117516 |
| RP11-282J12 | 3p12.3 | AC024222 | RP11-354H21 | 3p12.3 | AC117462 |
| RP11-983G16 | 3p12.3 | AC073856 | RP11-347G18 | 3p12.3 | AC112508 |
| RP11-582I9 | 3p12.3 | AC119043 | RP11-529F4 | 3q21.3 | AC080007 |
| RP11-274A21 | 3p12.3 | AC108746 | RP11-271G19 | 3q21.3 | AC055727 |
| RP11-705F7 | 3p12.3 | AC130003 | RP11-93K22 | 3q21.3 | AC083906 |
| RP11-642N14 | 3p12.3 | AC117481 | RP11-77P16 | 3q21.3 | AC130888 |
| RP11-241K7 | 3p12.3 | AC131233 | RP11-687B8 | 3q21.3 | AC112649 |
| RP11-803B1 | 3p12.3 | AC139453 | RP11-129J11 | 3q21.3 | AC093004 |
| CTD-2026G6 | 3p12.3 | AC133041 | RP11-202A13 | 3q22.1 | AC108742 |
| RP11-413E6 | 3p12.3 | AC108724 | RP11-217F11 | 3q23 | AC092988 |
| RP11-441E2 | 3p12.3 | AC133123 | RP11-25K24 | 3q25.1 | AC024886 |

## 2.7. Materials for promoter methylation analysis in human cell lines

5-Azadeoxycytidine (AZA) treatment of cell lines followed by RT-PCR analysis was performed on human Hela and HEK 293 cell lines. Medium and other components for cell culture are listed in table 2.12.

| ***Name*** | ***Tissue origin*** | ***Morphology*** |
|---|---|---|
| Hela | Human cervix | Epithelial |
| HEK 293 | Human kidney | Epithelial |

Table 2.12. Materials for cell culture and methylation analysis.

| ***Material*** | ***Company*** | ***Cat.No.*** |
|---|---|---|
| DMEM (without L-glutamine) | PAA Laboratories | E15-005 |
| Fetal bovine serum (FBS) | GIBCO (Invitrogen corporation) | 10270106 |
| L-glutamine (200 mM in 0.85% NaCl solution) | Cambrex Bio Science Verviers | BE17-605E |
| Antibiotic-antimycotic solution (Pen-Strep) (10,000 U Penicillin/ml, 10,000 µg Streptomycin/ml) | Cambrex Bio Science Verviers | DE17-602E |
| 1x trypsin-EDTA solution | GIBCO (Invitrogen corporation) | 25300-054 |
| Freezing medium-DMSO | GIBCO (Invitrogen corporation) | 11101-011 |
| DPBS (Dulbecco's phosphate buffered saline -0.0095M $PO_4$ without Ca and Mg) | Cambrex Bio Science Verviers | BE17-512F |
| 5'-aza-2'-deoxycytidine | Sigma-Aldrich | A3656 |

# 3.1. Microbiological techniques

## 3.1.1. Bacterial cultures

Bacteria can be cultivated on agar plates or in liquid medium.

- Culture on agar plates
  Streak the bacteria from glycerol stocks or LB agar stabs on selective agar plate (with selective antibiotic) and incubate at 37°C for 12-16 hours.
- Culture in liquid medium
  Inoculate a single colony from an agar plate in the corresponding selective LB liquid medium: 3 ml for mini-prep and 40 ml for mini-prep. Incubate at 37°C for 12-16 hours with shaking (~220 rpm).

## 3.1.2. Bacterial storage

- For short storage: Streaked plate can be stored at 4°C for one month.
- For long storage: Carefully mix 850 µl of overnight cultured bacteria with 150 µl of 100% glycerol and store at -80°C for years.

# 3.2. DNA and RNA isolation

## 3.2.1. Plasmid DNA isolation

The plasmid DNA isolation was done with the NucleoSpin® plasmid Kit from Macherey-Nagel. The bacterial pellets were resuspended and plasmid DNA was released from the *E.coli* host cells by SDS/alkaline lysis. The resulting lysate was neutralized by buffer A3, creating appropriate conditions for binding of plasmid DNA to the silica membrane of the provided column. The contaminations such as salts, metabolites and soluble macromolecular cellular components were removed by simple washing with buffer A4 which contains ethanol. Pure plasmid DNA was finally eluted under low ionic strength conditions with slightly alkaline buffer AE (5 mM Tris-Cl, pH8.5). This standard protocol is suited for plasmids of <15 kb length.

◆ ***Standard protocol***

- Cultivate and harvest bacterial cells: Centrifuge 2 ml of a saturated overnight *E. coli* LB culture for 30 sec at 11,000x g. Discard the supernatant.
- Cell lysis:
  - Add 250 µl buffer A1 (with RNase A). Vortex bigorously to resuspend the cell pellet.
  - Add 250 µl buffer A2. Mix gently by inverting the tube 6-8 times. Incubate at room temperature for <5 min.
  - Add 300 µl buffer A3. Mix gently by inverting the tube 6-8 times.
- Clarification of lysate: Centrifuge for 5-10 min at 11,000x g.
- Bind DNA: Load the supernatant from last step onto the column. Centrifuge for 1 min at 11,000x g. Discard flowthrough.
- Wash silica membrane:
  - Add 500 µl prewarmed buffer AW (50°C). Centrifuge for 1 min at 11,000x g.
  - Add 600 µl buffer A4 (with ethanol). Centrifuge for 1 min at 11,000x g. Discard flowthrough and repeat once.
- Dry slilica membrane: Insert the column into the 2 ml collecting tube. Centrifuge for 2 min at 11,000x g.
- Elute highly pure DNA: Place the column in a 1.5 ml microcentrifuge tube and add 50 µl buffer AE. Incubate for 1 min at room temperature. Centrifuge for 1 min at 11,000x g.

### 3.2.2. BAC DNA isolation

The genomic DNA isolation from large constructs (up to 250 kb) such as BACs (Bacterial Artificial Chromosomes) was done with the PhasePrep$^{TM}$ BAC DNA Kit from Sigma. *E.coli* culture was harvested by centrifugation and subjected to a modified alkaline-SDS lysis procedure. Nucleic acid was precipitated from the cleared lysate; residual RNA was removed by a short digestion at elevated temperature with an RNase cocktail. Endotoxins and other impurities were removed by simple temperature-mediated extraction and phase separation.

◆ ***Standard protocol for mini scale preparation***

- Harvest Cells: Centrifuge 30-40 ml of an overnight culture at 4,000-5,000x g for 10 min. Remove all supernatant.
- Resuspend cells: Add 2 ml of Resuspension Solution (with RNase A) to the bacterial cell pellet. Completely resuspend the pellet by vigorously vortexing.
- Lyse cells: Add 2 ml of Lysis Solution to resuspended cells. Mix immediately by gently inverting the tube 4-6 times and incubate at room temperature for 5 min.
- Neutralize: Add 2 ml of chilled Neutralization Solution to the lysate. Mix immediately by gently inverting the tube 6-8 times. Incubate the tube on ice for 5-10 min. Mix again during the ice incubation by gently inverting the tube 2-3 times. Centrifuge at 15,000x g for 20 min at 4°C to pellet the cell debris. Carefully transfer the cleared lysate (supernatant) to a clean centrifuge tube.
- Precipitate nucleic acids: Add 5 ml of room temperature isopropanol to the cleared lysate. Mix thoroughly by gentle inversion and centrifuge at 15,000x g for 20 min at 4°C. Pour off the supernatant and wash the pellet with 2 ml of 70% ethanol at room temperature. Centrifuge at 15,000x g for 5 min at 4°C and discard the supernatant. Pipette off the liquid and air-dry the pellet briefly (5-8 min).
- Digest residual RNA. Add 650 µl of Elution Solution to resuspend the pellet. Gently swirl and dissolve the DNA for 5 min. Add 1µl of the RNase Cocktail to the tube, mix briefly and incubate at 60°C for 10 min. Swirl 1-2 times during the digestion to dissolve any residual RNA.
- Adjust salt concentration: Centrifuge the tube for 1 min to collect the liquid. Transfer the solution to a 1.5 ml microcentrifuge tube. Add 50 µl of Sodium Acetate Buffer Solution (3M, pH7.0) and mix briefly.

- Remove endotoxins and other impurities:
  - Add 120 µl of Endotoxin Removal Solution and mix thoroughly by inversion for 30 sec. Chill the tube on ice for ≥ 5 min. Mix 1-2 times during the incubation on ice.
  - Incubate the tube for 5 min at 37°C. Centrifuge at ~16,000x g for 3 min at room temperature.
  - Carefully transfer the clear upper phase into a new 1.5 ml microcentrifuge tube and repeat the treatment with Endotoxin Removal Solution once.
- Precipitate DNA: Add 700 µl of room temperature DNA Precipitation Solution. Mix thoroughly by gentle inversion and centrifuge at maximum speed for 20 min at 4°C. Remove the supernatant and wash the pellet with 500 µl of 70% ethanol at room temperature. Centrifuge again for 5 min and remove the supernatant. Repeat the wash with 100 µl of 70% ethanol. Remove all of the supernatant and dry in a vacuum system for 10 min. Dissolve the DNA in 50 µl $H_2O$ by incubating the solution for ≥ 30 min at 65°C.

### 3.2.3. Gel extraction of PCR fragments

The PCR fragments were separated on agarose gel with optimal concentration (Table 3.1). Gel images were taken with an Intas Gel Jet Imager 2000. For DNA visualization, the gel was stained in 0.001% ethidium bromide solution for 10 mins and observed under U.V. light (366 nm). The band with expected product size was extracted from the gel with the NucleoSpin® Extract kit from Macherey-Nagel. DNA binds to a silica membrane in the presence of chaotropic salts (buffers NT1 and NT2). Buffer NT1 contains additional components in order to dissolve agarose gel slices. Afterwards, the dissolved agarose mixtures were loaded onto columns. Contaminations such as salts and soluble macromolecular components were removed by washing with ethanolic buffer NT3. Purified DNA was eluted under low ionic strength conditions with slightly alkaline buffer NE (5 mM Tris-Cl, pH8.5).

◆ ***Standard protocol for DNA extraction from agarose gel***

- Excise DNA fragment with a clean scalpel from agarose gel.
- Gel lysis: Add 300 µl NT1 buffer per 100 mg of agarose gel. Incubate sample at 50°C for 5-10 min until the gel slices are dissolved. Vortex the sample every 2-3 min to help to dissolve the gel slices.
- Bind DNA: Load the sample on a NucleoSpin® Extract column and centrifuge for 1 min at 8,000x g. Discard flowthrough.
- Wash silica membrane:
  - Add 500 µl buffer NT2. Centrifuge for 1 min at 11,000x g and discard flowthrough.
  - Add 600 µl buffer NT3. Centrifuge for 1 min at 11,000x g and discard flowthrough.
  - Add 200 µl buffer NT3. Centrifuge for 2 min at 11,000x g to remove buffer NT3 and dry the silica membrane.
- Elute DNA: Place the column into a clean 1.5 ml microcentrifuge tube. Add 25 µl prewarmed (70°C) elution buffer NE (for fragments >5-10 kb) and incubate at room temperature for 1 min. Centrifuge for 1 min at 11,000x g.

Table 3.1. Optimal agarose concentrations for resolving DNA fragment.

| ***Agarose Concentration*** **[%w/v]** | ***DNA size*** **[kb]** |
|---|---|
| 0.3 | 5-60 |
| 0.6 | 1-20 |
| 0.7 | 0.8 – 10 |
| 0.9 | 0.5 – 7 |
| 1.2 | 0.4 – 6 |
| 1.5 | 0.2 – 3 |
| 2.0 | 0.1 – 2 |

## 3.2.4. RNA isolation from tissue samples

TRIzol Reagent® was used to isolate total RNA from cell pellets and tissues. TRIzol is a ready-to-use mono-phasic solution of phenol and guanidine isothiocyanate. This method is an improvement of the single-step RNA isolation method of Chomczynski and Sacchi (1987). TRIzol maintains the integrity of RNA, while disrupting cells and

dissolving cell components during sample homogenization or lysis. Addition of chloroform followed by centrifugation separates the solution into an aqueous phase and an organic phase. RNA remains exclusively in the aqueous phase. After transfer of the aqueous phase, the RNA is recovered by precipitation with isopropanol.

- ***Standard protocol for RNA isolation from tissue samples***
  - Homogenization: Homogenize tissue sample in 1 ml of TRIzol Reagent® per 50-100 mg of tissue using a power homogenizer.
  - Phase separation: Incubate the homogenized sample for 5 min at room temperature. Add 200 µl chloroform for each milliliter of TRIzol. Shake tubes vigorously by hand for 15 sec and incubate at room temperature for 2-3 min. Centrifuge the sample at 12,000x g for 15 min at 2-8°C.
  - RNA precipitation: Transfer the aqueous phase to a fresh tube. Add 500 µl isopropanol to each milliliter TRIzol used for homogenization. Incubate sample at room temperature for 10 min and centrifuge at 12,000x g for 10 min at 2-8°C. Remove supernatant. Wash the RNA pellet once with 1 ml of 75% ethanol per 1 ml of TRIzol. Mix the sample by vortexing and centrifuge at 7,500x g for 5 min at 2-8°C.
  - Redissolving the RNA: Vacuum-dry RNA pellet for 5-10 min. Dissolve the RNA in RNase-free $H_2O$.

### 3.2.5. RNA isolation from cultured cell pellets

The NucleoSpin® RNA II kit was used to isolate RNA from cultured human cells. Cells were lysed by incubation in solution containing large amounts of chaotropic ions. This buffer immediately inactivates RNases and creates appropriate binding conditions for adsorption of RNA to the silica membrane. DNA contamination was removed by DNase I digestion of the DNA bound to the silica membrane. Simple washing steps with two different buffers were used to remove salts, metabolites and macromolecular cellular components. Purified RNA was eluted under low ionic strength conditions with RNase-free $H_2O$.

- ***Standard protocol for total RNA purification from cultured cells***
  - Homogenization of sample and cell lysis: Up to 5x $10^6$ eukaryotic cultured cells are collected by centrifugation and lysed by adding 350 µl buffer RA1 and 3.5 µl ß-mercaptoethanol to the cell pellet. Vortex vigorously.
  - Filtration of the lysate: Place NucleoSpin® Filter units in a collecting tube, apply the mixture and centrifuge for 1 min at 11,000x g.
  - Adjust RNA binding conditions: Discard the NucleoSpin® Filter unit and add 350 µl ethanol (70%) to the homogenized lysate. Mix by vortexing.
  - Bind RNA: Load the lysate onto a NucleoSpin® RNA II column placed in a 2 ml centrifuge tube. Centrifuge for 30 sec at 8,000x g. Place the column in a new collecting tube.
  - Desalt silica membrane: Add 350 µl MDB (Membrane Desalting Buffer) and centrifuge at 11,000x g for 1 min to dry the membrane.
  - Digest DNA:
    - Prepare DNase reaction mixture in a sterile microcentrifuge tube: for each isolation, add 10 µl reconstituted DNase I (provided with the kit) to 90 µl DNase reaction buffer. Mix by flicking the tube.
    - Apply 95 µl DNase reaction mixture directly onto the center of the silica membrane of the column. Incubate at room temperature for 15 min.
  - Wash and dry silica membrane:
    - First wash: Add 200 µl buffer RA2 to the NucleoSpin® RNA II column. Centrifuge for 30 sec at 8,000x g. Place the column into a new collecting tube.
    - Second wash: Add 600 µl buffer RA3 to the NucleoSpin® RNA II column. Centrifuge for 30 sec at 8,000x g. Discard flowthrough.
    - Third wash: Add 250 µl buffer RA3 to the NucleoSpin® RNA II column. Centrifuge for 2 min at 11,000x g. Place the column into a nuclease-free 1.5 ml microcentrifuge tube (supplied with the kit).
  - Elute highly pure RNA: Elute the RNA in 60 µl Rnase-free $H_2O$ (supplied with the kit) and centrifuge at 11,000x g for 1 min.

## 3.2.6. Photometric quantification of nucleic acid concentration

The nucleic acid (DNA and RNA) concentration and purity can be determined by spectrophotometric analysis. The maximum absorbance for nucleic acid is measured at

a wavelength of 260 nm.

1 $OD_{260}$ corresponds to: 50 µg/ml double strand DNA

40 µg/ml single strand DNA or RNA

20 µg/ml oligonucleotide

The DNA concentration (µg/ml) = $OD_{260}$ x dilution factor x 50

The RNA concentration (µg/ml) = $OD_{260}$ x dilution factor x 40

The measurements and calculations were performed using an Eppendorf Biophotometer.

By measuring the maximum absorbance for proteins at a wavelength of 280 nm, it is possible to evaluate the purity of sample by calculating the ratio of absorbance at 260 nm and 280 nm ($OD_{260}/OD_{280}$). The value should be between 1.8 and 2.0.

## 3.3. PCR (Polymerase Chain Reaction)

PCR is an enzymatic method for exponential amplification of specific DNA fragments *in vitro* (Saiki et al., 1985). It uses multiple cycles of template denaturation, primer annealing, and primer elongation (Saiki *et al.,* 1985). Since the amplified products from the previous cycle serve as templates for the next cycle, the amplification is an exponential process and a highly sensitive technique for nucleic acid detection. After 20-30 PCR cycles the amount of amplified DNA product is usually sufficient for visualization on an ethidium bromide-stained gel. PCRs were performed in an Eppendorf Mastercycler or an Eppendorf Mastercycler Gradient.

### 3.3.1. Standard PCR

The annealing temperature for standard PCR depends on the melting temperature of the primers used. Standard PCR was performed in total volume of 25 µl. The necessary reagent are listed in table 3.2.

Table 3.2. Reagents for standard PCR.

| | |
|---|---|
| 10x PCR buffer (without $MgCl_2$) | 2.5 µl |
| 50 mM $MgCl_2$ | 1.25 µl |
| 10 mM dNTPs | 0.5 µl |
| 10 µM forword primer | 1 µl |
| 10 µM reverse primer | 1 µl |
| 5 U/µl Taq Polymerase | 0.3 µl |
| DNA template | ~100 ng |
| $H_2O$ | add up to 25µl |

◆ ***Standard PCR program:***

1) First denaturation 94°C 3 min
2) Denaturation 94°C 30 sec
3) Annealing Variable * 30 sec
4) Elongation 72°C 30 sec#
5) Cycles Go to step 2) for 29 cycles
6) Final extension 72 °C 10 min

***** Primer specific temperature calculated by the formula (Maniatis *et al.,* 2001):

Annealing temperature (°C) = 4 x (G + C) + 2 x (A + T)

# If the predicted size of PCR is > 1kb, prolong elongation time 1 min per kb.

## 3.3.2. Expand Long Template PCR System

The system from Roche Applied System was used to amplify long (~10 kb) fragments from human BAC clone DNA. The products were purified by gel extraction and labeled by nick-translation for FISH. The protocol listed in table 3.3 is according to the user manual.

Table 3.3. Reagents for PCR amplification of ~10 kb DNA fragments.

| *Component* | | *Volume* | *Final concentration* |
|---|---|---|---|
| **Master Mix 1** | sterile bidist. $H_2O$ | | up to 25 µl |
| | 10 mM dNTP mix | 1.75 µl | 350 µM |
| | 10 µM forward primer | 1.5 µl | 300 nM |
| | 10 µM reverse primer | 1.5 µl | 300 nM |
| | template DNA | x µl | 100 ng BAC DNA |
| **Master Mix 2** | sterile bidist. $H_2O$ | | up to 25 µl |
| | 10x PCR buffer with $MgCl_2$ | 5 µl | buffer 1 (1.75 mM) |
| | enzyme mix | 0.75 µl | |

Pipet together master mix 1 and mix 2 and start the following cycling program.

◆ ***Expand Long Template PCR program for amplification of ~ 10 kb fragment***

| | | | |
|---|---|---|---|
| 1) | Denaturation template: | 94°C | 2 min |
| 2) | Denaturation: | 94°C | 10 sec |
| 3) | Annealing: | Variable * | 30 sec |
| 4) | Elongation: | 68°C | 8 min |
| 5) | Go to step 2), repeat 9 cycles | | |
| 6) | Denaturation: | 94°C | 10 sec |
| 7) | Annealing: | Variable * | 30 sec |
| 8) | Elongation: | 68°C | 8 min + cycle elongation of 20 sec for each cycle# |
| 9) | Go to step 6), repeat 19 cycles | | |
| 10) | Final elongation | 68°C | 7 min |

***** Primer specific temperature calculated by the formula (Maniatis *et al.,* 2001):

Annealing temperature (°C) = 4x (G + C) + x (A + T)

# e.g. cycle No.11 20 sec, cycle No.12 40 sec and cycle No. 13 60 sec longer.

### 3.3.3. DOP (Degenerate Oligonucleotide Primed) PCR

This method uses a single degenerate oligonucleotide primer (which is in fact a set of oligonucleotide sequences that have been synthesized in parallel, having the same base at certain nucleotide positions, while differing at other positions) to amplify a variety of related target DNAs. DOP PCR is widely used to amplify complex DNA probes from large-insert clones, i.e. BACs, PACs and YACs, flow-sorted or microdissected DNA libraries. In this work, DOP PCR was used to produce larger amounts of primate chromosome specific DNA libraries from pre-amplified primate flow-sorted chromosomal libraries (kindly provided by Prof. Malcolm Ferguson-Smith and Dr. Fengtang Yang, Cambridge University). The universal primer sequence is: 5'-CCGACTCGAGNNNNNNATGTGG-3' (Telenius *et al.,* 1992).

◆ ***Contents of DOP PCR reaction***

- 10x PCR buffer 5 µl
- 50 mM $MgCl_2$ 2 µl
- 10 mM dNTPs 2 µl
- 25 µM primer 2 µl
- 5 U/µl *Taq* 0.5 µl
- Template 2 µl ( pre-amplified flow-sorted chromosomal DNA)
- $H_2O$ up to 50 µl

◆ ***Standard program***

| | | | |
|---|---|---|---|
| 1) | First denaturation | 95°C | 5 min |
| 2) | Denaturation | 95°C | 1 min |
| 3) | Annealing | 62°C | 1.5 min |
| 4) | Elongation | 72°C | 3 min |
| 5) | Cycles | Go to step 2) for 29 cycles | |
| 6) | Final extension | 72 °C | 7 min |

### 3.3.4. RT (Reverse Transcription) PCR

cDNAs were synthesized from varying amounts of total RNA using the SuperScript™ III Reverse Transcriptase from Invitrogen. Reverse GSPs (gene-specific primers) were used to synthesize the first strand cDNA. Following first-strand synthesis the RNA complementary to cDNA was removed by digestion with RNase H to increase the

sensitivity of PCR. Target cDNA was then amplified with gene specific primer pairs by standard PCR. Only 2 µl (10%) of the first-strand reaction was used as template for subsequent PCR.

- ***Fisrt-strand cDNA synthesis***
  - Denaturation:
    - Mix samples as follows:
      1) 1 µl of 10 mM dNTP mix
      2) 1 µl of 2 µM reverse GSP
      3) 0.5-5 µg total RNA
      4) Add $H_2O$ up to 13 µl
    - Heat mixture at 65°C for 5 min, chill on ice for >1 min.
  - cDNA synthesis:
    - Add following reagents to the mixture
      1) 4 µl of 5x First-stand buffer
      2) 1 µl 0.1 M DTT
      3) 1 µl RNaseOUT™ Recombinant RNase inhibitor
      4) 1 µl SuperScript™ III RT
    - Mix by pipetting gently up and down.
    - Incubate at 55°C for 1 hr.
  - Inactivation of the reaction and removal of RNA:
    - Incubate at 70°C for 15 min to inactivate the reaction.
    - Add 0.7 µl RNase H (5 U/µl) and incubate at 37°C for 20 min to remove RNA.

## 3.4. 5'-RACE (rapid amplification of cDNA ends)

The BD SMART™ RACE cDNA amplification kit uses the SMART (Switching Mechanism At 5'End of RNA Transcript) cDNA synthesis technology which provides a mechanism for generating full-length cDNAs in reverse transcription reactions (Zhu *et al.,* 2001). BD PowerScript reverse transcriptase is a variant of Moloney Murine Leukemia Virus Rreverse transcriptase that exhibits terminal transferase activity and adds 3-5 residues (predominantly dC) to the 3' end of first-strand cDNA when it reaches the end of the RNA template (Figure 3.1). The BD SMART oligo contains a terminal stretch of G residues that anneals to this dC-rich cDNA tail and serves as an extended template for RT. BD PowerScript RT switches templates from the mRNA

molecule to the BD SMART oligo and thus generates a complete cDNA copy of the original RNA with the additional BD SMART oligo sequence at the end. Since the transferase activity of RT is most efficient at the end of the RNA template, the BD SMART oligo sequence is typically added to complete first-strand cDNAs. The RACE reaction was performed in an Eppendorf Mastercycler. The amplified product was cloned into TOPO® vector for subsequent sequencing analysis. The sequences of oligonucleotides provided with the kit and the components for 5'RACE are listed in tables 3.4 and 3.5.

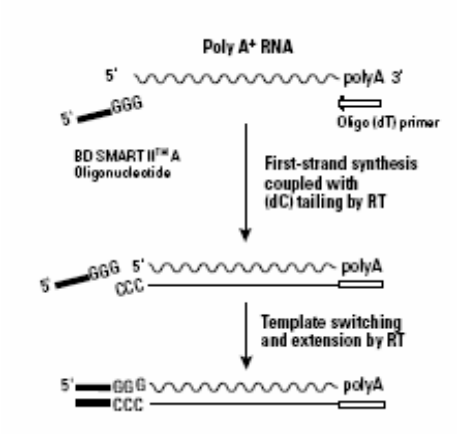



Figure 3.1. Mechanism of BD SMART™ cDNA synthesis. First-strand synthesis is primed with a modified oligo(dT) primer. When reaching the end of the mRNA template, reverse transcriptase adds several dC residues. The BD SMART II A oligonucleotide anneals to the tail of the cDNA and serves as an extended template for BD PowerScript RT. (From BD Biosciences Clontech protocol No. PT3269-1)

Table 3.4. Oligonucleotides for 5'RACE (provided with the kit).

| ***Oligonucleotide*** | ***Sequence*** |
|---|---|
| SMART II A oligonucleotide | 5'-AAGCAGTGGTATCAACGCAGAGTACGCGGG-3' |
| 5'-RACE CDS primer | 5'-$(T)_{25}N_{-1}N$-3' (N=A,C,G, orT; $N_{-1}$=A,G, or C) |
| 10x Universal Primer A Mix (UPM) | Long (0.4 μM):<br>5'-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3'<br>Short (2 μM):<br>5'-CTAATACGACTCACTATAGGGC-3' |

◆ ***Standard protocol for first-strand cDNA synthesis***

- Set up reaction as follows:
  - 1-3 µl (50 ng to 1 µg) of total RNA sample
  - 1 µl 5'-CDS primer*
  - 1 µl BD SMART II A oligo*
  - sterile $H_2O$* to a final volume of 5µl
- Mix and spin the tube briefly.
- Incubate the tube at 70°C for 2 min.
- Chill the tube on ice for 2 min.
- Spin the tube briefly to collect the contents at the bottom.
- Add the following to the reaction tube:
  - 2 µl of 5x first-strand buffer*
  - 1 µl DTT (20 mM)*
  - 1 µl dNTP Mix (10 mM)*
  - 1 µl BD PowerScript Reverse Transcriptase*
- Mix the contents by gentle pipetting.
- Spin the tube briefly to collect the contents at the bottom.
- Incubate the tubes at 42°C for 1.5 hrs.
- Dilute the first-strand reaction product with Tricine-EDTA buffer*:
  - Add 20 µl if template is <200 ng of total RNA.
  - Add 100 µl if template is >200 ng of total RNA.
- Heat tubes at 72 °C for 7 min.
- Ready to go to RACE PCR or store sample at -80°C.

(*: materials are provided with the BD SMART$^{TM}$ RACE cDNA amplification kit)

◆ ***Standard protocol for 5'-RACE***

- Set up PCR Master Mix as follows (for one reaction):
  - 34.5 µl PCR-grade $H_2O$†
  - 5 µl 10x BD Advantage 2 PCR buffer†
  - 1 µl dNTP Mix (10 mM) †
  - 1 µl 50x BD Advantage 2 Polymerase Mix†
- Add the components in the order shown in table 3.5 and mix gently.
- *Start standard program*

1) 5 cycles:
   94°C 5 sec
   72°C 3 min
2) 5 cycles:
   94°C 5 sec
   70°C 10 sec
   72°C 3 min
3) 30 cycles:
   94°C 5 sec
   68°C 10 sec
   72°C 3 min

Table 3.5. Reagents for 5'-RACE.

| ***Component*** | ***1 5'-RACE Sample*** | ***2 5'-TFR* (+ Control)*** | ***3 UPM* only (- Control)*** | ***4 GSP* only (- Control)*** |
|---|---|---|---|---|
| 5'-RACE-Ready cDNA | 2.5 µl | 2.5 µl | 2.5 µl | 2.5 µl |
| UPM (10x) † | 5 µl | 5 µl | 5 µl | / |
| GSP (10 µM) | 1 µl | / | / | 1 µl |
| Control 5'-RACE TFR primer (10 µM) † | / | 1 µl | / | / |
| $H_2O$† | / | / | 1 µl | 5 µl |
| Master Mix | 41.5 µl | 41.5 µl | 41.5 µl | 41.5 µl |
| Final volume | 50 µl | 50 µl | 50 µl | 50 µl |

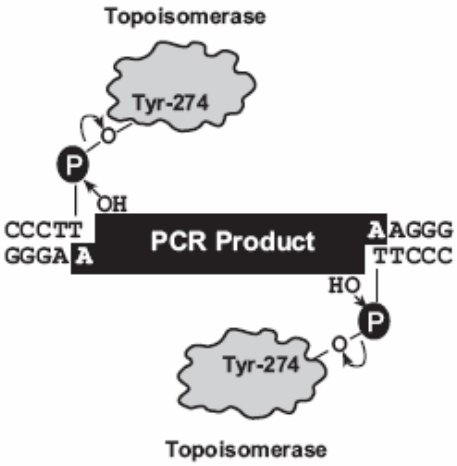
* 5'RACE with TFR primer will amplify the 5'end of the transferring receptor (TFR) cDNA. Skip this reaction if the RNA is not human. UPM: Universal Primer A mix. GSP: Gene Specific Primer

† Materials are provided with the BD Advantage 2 PCR kit.

# 3.5. Cloning of PCR products

## 3.5.1. TOPO® TA cloning

TOPO® TA cloning provides a highly efficient one-step cloning strategy for the direct insertion of *Taq* polymerase-amplified PCR products into a plasmid vector.




Plasmid vectors pCR®II-TOPO® and pCR®2.1- TOPO® are supplied in a linearized form with a single 3'-thymidine (T) overhang and topoisomerase I covalently bound to the vector. *Taq* polymerase has a nontemplate-dependent terminal transferase activity that adds a single deoxyadenosine (A) to the 3' ends of PCR products. This allows efficient ligation of PCR products into the vector. Topoisomerase I binds to duplex DNA at specific sites and cleaves the phosphodiester backbone after 5'-CCCTT in one strand. The energy from the broken phosphodiester backbone is conserved by formation of a covalent bond between the 3' phosphate of the cleaved stand and a tyrosyl residue (Tyr-274) of topoisomerase I. The phosphotyrosyl bond between DNA and enzyme can subsequently be attacked by the 5' hydroxyl of the original cleaved strand, reversing the reaction and releasing topoisomerase. The PCR products to be cloned should have a final extension step for at least 10 min to ensure that all PCR products are full length and 3' adenylated. The cloning site on the vector (pCR®II-TOPO® or pCR®2.1- TOPO®) lies within the *lacZα* gene which codes for beta galactosidase. The ligation of PCR product disrupts the *lacZα* gene so that the substrate of beta-galactosidase, X-gal, can not be converted into an insoluble blue dye. This allows blue/white screening for the selection of insert-containing clones. The materials needed for cloning are listed in table 3.6.

Table 3.6. Reagents necessary for TOPO® TA cloning.

| | |
|---|---|
| Components provided with the kit | TOPO® vector |
| | Salt solution |
| | One Shot® TOPO10 *E.coli* competent cells |
| | S.O.C. medium |
| Other materials | Fresh PCR products |
| | Selective LB plates containing 50 μg/ml kanamycin and 40 μl of 40 mg/ml X-Gal spread on surfaces |

◆ ***Standard protocol for TOPO® TA cloning***

- PCR products: fresh PCR products are purified by gel extraction if more than one band are amplified.
- TOPO® TA cloning reaction
  - Set up reaction as follows:
    1) $H_2O$ up to final volume of 6 μl
    2) 0.5-4 μl fresh PCR products
    3) 1 μl salt solution
    4) 1 μl TOPO® vector
  - Mix gently and incubate at room temperature for 5-30 min.
- One Shot® chemical transformation
  - Thaw One Shot® *E.coli* competent cells on ice for 5 min.
  - Add 2-6 μl of cloning reaction to a vial of One Shot® TOPO10 *E. coli* competent cells and mix gently.
  - Incubate on ice for 30 min.
  - Heat-shock at 42°C for 30 sec.
- Recovery and plating
  - Add 250 μl of S.O.C. medium at room temperature and recover the bacteria by shaking horizontally at 220 rpm for 1 hr at 37°C.
  - Spread 100-150 μl bacterial suspension on a selective plate.
  - Incubate 12-16 hrs at 37°C.

- Analysis of positive clones
  - Colony PCR: Pick up white clones. Replicate them on a new selective plate (in order to identify positive clones based on colony PCR results) and also in 50 µl TE. Use 10 µl of this TE as template for a standard PCR with GSP or M13 primers.
  - Sequencing: Culture 10 colonies showing bands of the predicted size after colony PCR with selective LB medium. Isolate plasmid DNA with NucleoSpin® plasmid kit. The sequence of each clone is determined as described in chapter 3.6.
  - Positive colonies are stored as glycerol stock at -80°C.

### 3.5.2. TOPO® TA XL cloning

This cloning method provides a highly efficient strategy for the cloning of long PCR products (3-10 kb). The plasmid vector pCR®-XL-TOPO® also contains a single 3'-thymidine (T) overhang and covalently bound topoisomerase I. The vector contains the lethal *E.coli* gene *ccd*B. This gene is fused to the C-terminus of the *LacZα* fragment. Ligation of a long PCR product disrupts expression of the *lacZα-ccd*B fusion gene permitting only the growth of insert-containing recombinants upon transformation. Because cells with non-recombinant vector are killed upon plating, blue/white screening is not required.

◆ ***Standard protocol for TOPO® TA XL cloning***

- PCR products: fresh PCR products are purified by gel extraction.
- TOPO® TA XL cloning reaction
  - Set up reaction as follows:
    1) $H_2O$ up to final volume of 6µl
    2) 0.5-4 µl fresh PCR products
    3) 1µl pCR®-XL-TOPO® vector
  - Mix gently and incubate at room temperature for 5 min.
  - Add 1µl of the 6x TOPO® Cloning Stop Solution (provided with the kit) and mix for several seconds at room temperature.
  - Briefly centrifuge the tube and place on ice.

- One Shot® chemical transformation
  - Thaw One Shot® *E.coli* competent cells on ice for 5 min.
  - Add 2-6 µl of cloning reaction to a vial of One Shot® TOPO10 *E. coli* competent cells and mix gently.
  - Incubate on ice for 30 min.
  - Heat-shock at 42°C for 30 sec.
  - Immediately chill the tube on ice and incubate for 2 min.
- Recovery and plating: As desicribed above in "standard protocol for TOPO® TA cloning".
- Analysis of positive clones: Pick up clones and analysis as desciribed above in "standard protocol for TOPO® TA cloning".

## 3.6. DNA sequencing

DNA sequencing was performed with CEQ 2000 Dye Terminator Cycle Sequencing (DTCS) Quick Start Kit and analyzed on CEQ™ 8000 Genetic Analysis System from Beckman Coulter ™.

◆ ***Standard protocol for plasmid DNA sequencing***

- Set up sequencing reaction as follows:
  1) $H_2O$ to adjust the final total volume to 20 µl.
  2) DNA template 300 ng.
  3) Denature the template at 96°C for 3 min, chill on ice for at least 2 min.
  4) Add 0.5 µl 10 µM primer.
  5) Add 4 µl DTCS Quick Start Master Mix.
- Thermal cycling program:
  1) 96°C 20 sec
  2) 50°C 20 sec
  3) 60°C 4 min
  4) Go to step 1) and repeat 29 cycles
  5) 4°C Hold.
- Ethanol precipitation:
  - Prepare 4 µl Stop Solution: 2 µl 3 M NaOAC + 2µl 0.1 M EDTA for each sample.
  - Transfer the sequencing reaction to Stop Solution.
  - Add 60 µl 100% ethanol (-20°C) and mix thoroughly. Centrifuge at 14,000 rpm at 4°C for 15 min. Carefully pipette off supernatant.
  - Rinse the pellet 2x with 180 µl 70% ethanol (-20°C). Centrifuge at 14,000 rpm at 4°C for 5 min and carefully pipette off supernatant.
  - Vacuum dry for 10 min.
  - Resuspend the sample in 30 µl of the SLS (Sample Loading Solution) (provided with the kit). The sample can be kept in SLS at -20°C for up to one week.
- Sample preparation for loading onto the sequencer:
  - Transfer the resuspended samples to the appropriate wells of the CEQ sample plate.
  - Overlay each of the resuspended samples with one drop of light mineral oil (provided with the kit).
  - Load the sample plate onto the CEQ and start the desired method.

## 3.7. Fluorescence *in situ* Hybridization

Fluorescence *in situ* hybridization (FISH) is a technique used to directly visualize DNA sequences on morphologically preserved cytological specimens such as metaphase chromosomes and interphase cell nuclei. Indirect FISH uses hapten-labeled DNA probes in combination with immunocytochemistry (secondary detection reagents). Both probe and target DNA must be denatured. In the hybridization reaction, complementary single-stranded sequences in the probe and chromosomal target are allowed to reanneal. After posthybridization washing and immunocytochemical detection of *in situ* bound hapten molecules, a specific fluorescent signal is produced at the hybridization site, which can be viewed by eye through the epifluroescence microscope.

### 3.7.1. Labeling of FISH probes by nick translation

The nick translation procedure was originally described by Rigby et al. (1977) and used for incorporating nucleotide analogs (Langer et al. 1981). The procedure incorporates one modified nucleotide (Digoxigenin-dUTP, Biotin-dUTP) at approximately every 20-25th position in the newly synthesized DNA. For *in situ* hybridization procedures, the length of the labeled fragments obtained from nick translation should be about 100-500 bases. The reaction was carried out in 50 µl of final solution as described below. The maximum reaction volume is 200 µl for labeling ~ 4 µg DNA.

◆ ***Nick translation***

- Reaction mix:
  - X µl $H_2O$ up to the final total volume of 50 µl
  - 1 µg BAC DNA (or gel extracted long range PCR products)
  - 5 µl 10x nicktranslation mixture
  - 5 µl 0.1 M ß-mecaptomethanol
  - 5 µl dNTP mix
  - 1 µl DIG or Biotin labelled dUTP (1 mM)
  - 3 µl 1/3000 DNase I dilution
  - 1 µl DNA polymerase
- Reaction is performed at 15°C for 2 hours.
- Chill the reaction tube on ice.
- Take a 7 µl aliquot from the tube and analyze it on a 1.5-2% agarose gel along with Gene Ruler™ 100bp DNA ladder.
- If the probe is between 100-500 base long, the reaction can be stopped. If the probe is longer than 500 base, add another 3 µl of 1/3000 DNase I dilution and incubate the reaction at 15°C for 1 hour. Analyze the size again on agarose gel.
- Stop the reaction by adding equal volume of stop mix.

◆ ***Purification of the labeled probe***

- Prepare Sephadex G-50 column: Block the bottom of one 1 ml syringe with cotton and fill up the tube with equilibrated Sephadex G-50, centrifuge at 1800 rpm for 2 min.
- Load mixed reaction sample and centrifuge at 1800 rpm for 5 min. The purified labeled probe can be stored at -20°C for more than one year.

## 3.7.2. Metaphase preparation and slide pretreatment

Metaphases were prepared from lymphoblastoid cell lines. Cells were arrested at metaphase by adding the spindle poison colcemid shortly before harvesting. Harvested cells were swollen in hypotonic solution (50 mM KCl at 37°C for 5-10 min) and then fixed in a 3:1 mixture of ice-cold methanol-acetic acid. To prepare slides, one or two drops of chromosome suspension were spread onto a very clean glass slide. For FISH, slides should be between one day and one week old. Extra slides can be stored in a

desiccating box for several months at -20°C.

◆ ***Pretreatment of slides***

- Equilibrate slides shortly in 2x SSC.
- RNase-treatment: Dilute RNase stock (20 mg/ml) 1:200 and incubate slides with 150 µl of diluted RNase solution under a coverslip for 30 min at 37°C.
- Wash slides 3x 5 min with 2x SSC.
- Incubate with 0.005% pepsin in 10 mM HCl for 10 min at 37°C.
- Wash slides 2x 5 min with 1x PBS and1x 5 min with 1x PBS/$MgCl_2$
- Fix slide for 10 min at room temperature with 1% formaldehyde in PBS/$MgCl_2$.
- Wash slides 1x 5 min with 1x PBS.
- Dehydrate slides in an ethanol series (70%, 80%, 99%) and then air-dry them.

### 3.7.3. Denaturation and hybridization

To prevent unspecific hybridization, repetitive DNA sequences in the probe DNA were blocked by prehybridization with unlabeled repetitive Cot-1 DNA.

◆ ***Denaturation of slides***

- Denature the slides with 150 µl of FA/SSC (70% formamid in 2x SSC) under coverslip at 73°C for 90 sec.
- Dehydrate slides in ethanol series as step in “Pretreatment of slides”.

◆ ***Precipitation and denaturation of probes***

- For probe DNA precipitation, add 50 µg of human Cot-1 DNA, 5 µg of herring sperm DNA, 5 µl of 3 M sodium acetate (pH5.2), and 2.5 volumes of 100% ethanol to the labeled probe (NT mixture).
- Incubate at -20°C overnight or at -80°C for 30 min.
- Centrifuge for 30 min at 4°C at 13,000x g.
- Discard the supernatant and add 500 µl of 70% ethanol to wash pellets. Centrifuge for 15 min at 4°C at 13,000 x g.
- Discard the supernatant and dry the pellet in a vacuum system for 10 min.
- Dissolve the pellet in 5 µl of deionized formamide for 30 min at 37°C.
- Add an equal volume of 2x hybridization mixture and incubate for at least 15 min at 37°C.
- Denature the probe for 10 min at 80°C, chill on ice, and allow annealing of the repetitive elements at 37°C for 30 min.

◆ ***Hybridization***

- Hybridize the chromosomes at 37°C overnight with 10 µl of the pre-annealed probe under a sealed coverslip.

## 3.7.4. Signal detection

The protocol described below allows simultaneous detection of biotinylated and digoxigenated probes.

◆ ***Detection***

- Wash slides 3x 5 min with 2x SSC at 37-45°C and 3x 5 min with 0.2x SSC at 55-65°C.
- Equilibrate the slides shortly with 4x SSC/Tween at 45°C.
- Incubate the slides with 150 µl blocking solution under coverslip for 30 min at 37°C. Wash slides shortly with 4 x SSC/Tween at 45°C.
- Centrifuge FITC-conjugated avidin for 2 min at 13,000 rpm. Dilute the supernatant 1:200 in antibody-dilution solution. Incubate the slide with 150 µl of avidin solution under coverslip for 30 min at 37°C. (Skip this and next step if only DIG-labeled probe is detected)
- Wash slides 3x 5 min with 4 x SSC/Tween at 45°C.
- During the third wash, centrifuge anti-avidin (skip if only DIG-labeled probe were used) and mouse anti-DIG antibody for 2 min at 13,000 rpm. Dilute anti-avidin 1:200 and mouse-anti-DIG 1:500 in antibody-dilution solution. Incubate the slide with 150 µl of antibody dilution under coverslip for 45 min at 37°C.
- Wash slides 3x 5 min with 4 x SSC/Tween at 45°C.
- During the third wash, centrifuge FITC-conjugated avidin (skip if only DIG-labeled probe is detected) and Cy3-conjugated sheep anti-mouse antibody for 2 min at 13,000 rpm. Dilute FITC/avidin 1:200 and Cy3-conjugated anti-mouse 1:500 in antibody-dilution solution. Incubate the slide with 150 µl of antibody solution under coverslip for 30 min at 37°C.
- Wash slides 3x 5 min with 4 x SSC/Tween at 45°C.
- Wash slides 3x shortly in $H_2O$ and air-dry in dark.
- Counterstain and mount slides with 20-30 µl of DAPI/Vectorshield.

***Microscope analysis and digital imaging***

Images were taken with a Zeiss epifluorescence microscope equipped with a CCD (charged coupled device) camera, which is controlled by an Apple Macintosh computer. Switching between different fluorochromes was achieved by changing filter sets. The spectral properties of fluorophores are shown in table 3.7. Macprobe imaging software was used to capture gray scale images and to superimpose the source images into a color image.

Table 3.7. Spectral properties of fluorophores used in FISH analysis. (From DIG Application Manual for nonradioactive *in situ* hybridization, 3rd edition, Roche Applied Science).

| | ***Color*** | ***Excitation max. (nm)*** | ***Emission max. (nm)*** |
|---|---|---|---|
| Fluorescein | Green | 494 | 523 |
| CY3 | Red | 552 | 565 |

## 3.8. Cell culture

### 3.8.1. Culture of human cell lines

Cell lines Hela and HEK 293 grow as adherent monolayers. Routinely, cells were maintained at 37°C in a humidified incubator (Binder CB150) with 5% $CO_2$ atmosphere. The DMEM culture medium was supplemented with 10% FBS, 1% L-glutamine and 1% pen-strep solution. Some key parameters, e.g. the volumes of growth medium per falsk are given below in table 3.8.

Table 3.8. Key parameters for cell culture.

| ***Flask*** | ***Surface area ($mm^2$)*** | ***Seeding density*** | ***Cells at confluency*** | ***Trypsin/EDTA (ml)*** | ***Growth medium (ml)*** |
|---|---|---|---|---|---|
| T-25 | 2,500 | $0.7x\ 10^6$ | $2.8x\ 10^6$ | 3 | 3-5 |
| T-75 | 7,500 | $2.1x\ 10^6$ | $8.4x\ 10^6$ | 5 | 8-15 |

### 3.8.2. Subculture of adherent human cell lines

Adherent cell lines will grow *in vitro* until they have covered the surface area available or the medium is depleted of nutrients. At this point the cell lines should be sub-cultured in order to prevent the culture dying.

◆ ***Standard procedure***

- View cultures using an inverted microscope (Axiovert 25 from Carl-Zeiss) to assess the degree of confluency and confirm the absence of bacterial and fungal contaminants.
- Remove old medium.
- Wash the cell monolayer with DPBS using half the volume of culture medium.
- Pipette trypsin/EDTA onto the washed cell monolayer using 1 ml per 25 $cm^2$ of surface area. Rotate flask to cover the monolayer with trypsin.
- Return flask to the incubator and leave for 2-10 min.
- Examine the cells under the microscope to ensure that all the cells are detached and floating. The side of the flasks may be gently tapped to release any remaining attached cells.
- Resuspend the cells in a small volume of fresh serum-containing medium to inactivate the trypsin.
- Transfer the cell suspension to a centrifuge tube and centrifuge at 150x g for 5 min.
- Resuspend the cells in 1-2 ml fresh serum-containing medium and transfer the required number of cells to a new labeled flask containing pre-warmed medium.
- Incubate as appropriate for the cell line.

### 3.8.3. Cryopreservation and resuscitation of frozen human cell lines

◆ ***Standard procedure for cyroperservation***

- Remove cells from the culture as described above.
- Resuspend the cell pellets in a small volume (1-2ml) of freezing medium.
- Pipette 1 ml aliquots of cells into cyroprotective ampoules that have been labeled with the cell lines' names, passage number and date.
- Place ampoules inside a passive freezer (Nalgene™ Cryo 1°C Freezing-Container). Fill freezer with isopropanol and place at -80°C overnight.
- Frozen ampoules should be transferred to a liquid nitrogen storage vessel.

- ***Standard procedure for resuscitation***
  - Prepare the flasks as appropriate and label with cell line's name, passage number and date.
  - Collect ampoule of cells from liquid nitrogen storage and place in a waterbath at 37°C. Submerge only the lower half of the ampoule. Allow the cells to thaw until a small amount of ice remains in the vial – usually 1-2 min.
  - Wipe the outside of the ampoule with a tissue moistened with 70% alcohol and hold tissue over ampoule to loosen lid.
  - Slowly, dropwise, pipette cells into pre-warmed growth medium to dilute out the DMSO in the freezing medium.
  - Incubate at the appropriate condition.

### 3.8.4. Methylation analysis of human cell lines

DNA methylation is an important regulatory mechanism of gene transcription. 5'-azacytidine and 5'-azadeoxycytidine are potent growth inhibitors and cytotoxic agents. They act as a demethylation agents by inhibiting DNA methyltransferase (Broday 1999). 5'-aza-2'-deoxcytidine was dissolved in 100% ethanol at a concentration of 1 mM and stored as stock solution at -20°C. Working solution were always prepared freshly before changing the medium: 1 µM (5 ml growth medium + 5 µl stock solution) and 10 µM (5 ml growth medium + 50 µl stock solution). The culture medium containing 5'-aza-deoxycytidine was changed every 12 hrs. Cell pellets were collected at 48 hr after 5'-aza-deoxycytidine treatment. RNA was isolated with Nucleospin® RNA II kit. The transcription level of target genes was analyzed by RT-PCR.

## 3.9. *In silico* sequence analysis

### 3.9.1. Databanks

- Databank from National Center for Biotechnology Information (NCBI)
  http://www.ncbi.nlm.nih.gov
- Databank from Wellcome Trust Sanger Institute (ensembl)
  http://www.ensembl.org
- The UCSC Genome Browser
  http://genome.cse.ucsc.edu
- The Human Genome Segmental Duplication Database

http://projects.tcag.ca/humandup/

This database was used to analyze the segmental duplications.

- The orangutan EST database

  http://mipc2.gsf.de/proj/cDNA/pest

### 3.9.2. Software for sequence analysis

- BLAST

  http://www.ncbi.nlm.nih.gov/BLAST/

  The BLAST 2 program with MegaBLAST was used to determine sequence similarity between BACs. Gap initiation and gap extension parameters were adjusted to minimize breaking long sequence matches into pieces at insertion/deletion sites. Percent identities of paralogous blocks were calculated from BLAST 2 outputs without considering insertions and deletions.

- REPEATMASKER

  at http://menu.hgmp.mrc.ac.uk/menu-bin/Nix/Nix.pl

  This program was used to analyze repetitive elements.

- GrailEXP

  at http://menu.hgmp.mrc.ac.uk/menu-bin/Nix/Nix.pl

  This program was used to search for CpG islands.

- HUSAR (Heidelberg Unix Sequence Analysis Resources)

  http://genome-dkfz-heidelberg.de

  http://genius.embnet.dkfz-heidelberg.de/menu/w2h/w2hdkfz/index.html

  "MultAlign" was used to comparing homologies among sequences.

  "Map" was used to translate DNA sequences into amino acid sequences.

- SignalP3.0 (Nielsen *et al.,* 1997 and Bendtsen *et al.,* 2004)

  http://www.cbs.dtu.dk/services/SignalP/

  This program was used to predict the presence and location of signal peptide cleavage sites.

# 4.1. FISH mapping of three evolutionary breakpoints between human chromosome 3 and orangutan chromosome 2

A previous reconstruction of the human (HSA, *Homo sapiens*) chromosome 3 rearrangements during primate evolution based on the hybridization of approximately 100 large-insert clones from human chromosome 3 to chromosomes of great apes, gibbon, Old and New World monkeys, demonstrated four large snytenic blocks that have been conserved in their entirety between human and Bornean orangutan (Figure 1.4). The three breakpoints between human 3p25.1, 3p12.3, 3q21.3 and Bornean orangutan (*Pongo pygmaeus*, PPY) chromosome 2 were mapped to 1-2 Mb intervals (Tsend-Ayush *et al.*, 2004). In order to narrow down these breakpoints further, STS markers in the breakpoint-spanning or breakpoint-flanking YACs were used to anchor the breakpoint regions in the HSA 3 genomic sequence. BAC contigs of the breakpoint regions were then assembled from the database and used as probes for comparative FISH analysis.

## 4.1.1. Evolutionary breakpoint between HSA 3p25.1 and PPY 2

The evolutionary breakpoint between the red and green segments (Figure 4.1A) was localized in YAC 772D2 at 15.3 Mb in HSA 3p25.1 (Tsend-Ayush *et al.*, 2004). The YAC-specific STS markers D3s3613 and WI-3852 were found in BAC clones RP11-321I9 and RP11-316A10, respectively. A contig containing these two BACs (Figure 4.1B) was assembled from the Wellcome Trust Sanger Institute, Ensembl contigs (http://www.ensembl.org) and UCSC genome browser (http://genome.cse.ucsc.edu/cgi-bin/hgGateway). FISH with individual BACs was then performed on human and Bornean x Sumatran orangutan hybrid chromosomes (Figure 4.2). According to the previous reconstruction, BACs of this contig should hybridize to the tip of the red arrow on the Bornean PPY 2q and Sumatran 2p subtelomere (indicated by a white triangle in Figures 4.1A & 4.2) and/or to the end of the green arrow on PPY 2q (indicated by a white star in 4.1A & 4.2).

FISH with RP11-135D4 and RP11-316A10 on orangutan chromosomes showed that these two BACs hybridized to Bornean PPY 2q and Sumatran PPY 2p telomere (Figures 4.2A and 4.2B). These sites correspond to the tip of the red arrow and are distal to the evolutionary breakpoint. RP11-421B21 and RP11-321I9 hybridized to distal PPY 2q (Figures 4.2E and 4.2F), which corresponds to the end of the green arrow, proximal to the breakpoint. RP11-72K12 and RP11-616M11 produced split signals on PPY 2 (Figures 4.2C and 4.2D) an, thus, contain the evolutionary breakpoint.

Collectively, these FISH mapping results suggest that RP11-135D4 and RP11-421B21 must be very close on the distal and proximal side of the inversion breakpoint, respectively. Although these two clones did not produce split FISH signals, the breakpoint could still be located within the last 10 kb of the BAC sequence but the split signals are undetectable by FISH. This narrows down the position of the 3p25.1 breakpoint to a 28.7 kb interval, corresponding to the segment from 56592 bp to 85332 bp of RP11-616M11 (AC090954).

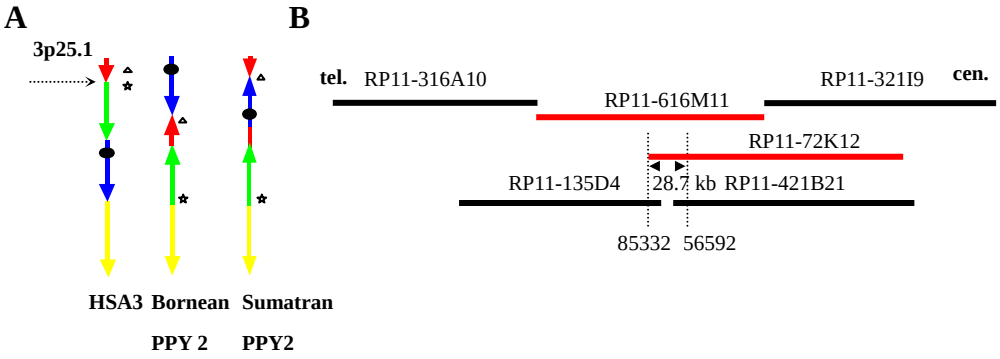



Figure 4.1. (A) Conservation of four syntenic segments (indicated by different colors) between Bornean PPY 2 and HSA 3. Three evolutionary breakpoints syntenic to HSA 3p25.1, 3p12.3 and 3q21.3 cause the different arrangement of these four segments in Bornean PPY 2 and HSA 3. The distal side of the 3p25.1 breakpoint region on PPY 2 is marked by a triangle (△).The proximal side of the 3p25.1 breakpoint region is indicated by a star (✶). (B) BAC contig of the HSA 3p25.1 breakpoint region: Each line represents a BAC clone with “finished” sequence. BAC names are written on the top. Red lines represent the breakpoint spanning clones (by FISH). The breakpoint is located in the marked 28.7 kb interval of BAC RP11-616M11 (56592-85332 bp).

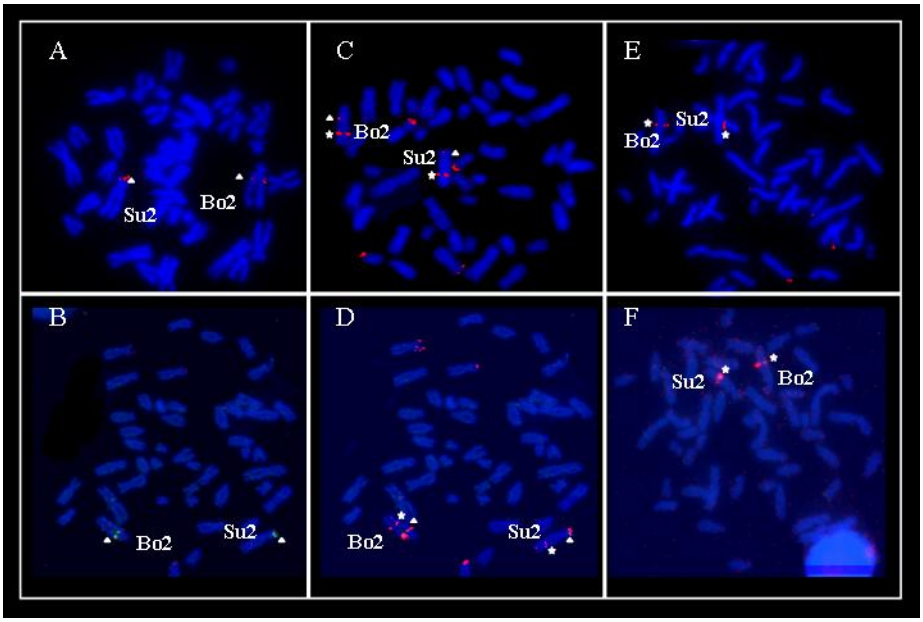



Figure 4.2. FISH with BAC clones of the 3p25.1 contig on orangutan hybrid metaphase spreads. Bo 2 and Su 2 indicate the Bornean and Sumatran PPY 2 chromosome forms. The distal side of the 3p25.1 breakpoint region on PPY 2 is marked by a triangle (△). The proximal side of the 3p25.1 breakpoint region is indicated by a star (★). (A, B) BACs RP11-135D4 and RP11-316A10 produce a single-copy hybridization signal distal to the 3p25.1 breakpoint. (C, D) The breakpoint-containing BACs RP11-72K12 and RP11-616M11 produce split signals on PPY 2. (E, F) BACs RP11-421B21 and RP11-321I9 produce single-copy hybridization signals proximal to the 3p25.1 breakpoint.

### 4.1.2. Evolutionary breakpoint between HSA 3p12.3 and PPY 2

This evolutionary breakpoint was localized in YAC 905D2 at 75 Mb in HSA 3p12.3. The YAC-specific STS markers WI-9886 and WI-2033 were located in BAC clones RP11-642N14 and RP11-58L10, respectively. A contig (Figure 4.3B) containing these two BACs was assembled from database and FISH with individual BACs was performed on human and orangutan chromosomes. According to a previous reconstruction, BACs of this contig should hybridize to the tip of the green arrow on PPY 2q (indicated by a dot in Figures 4.3A and 4.4) and/or to the end of the blue arrow at the Bornean PPY 2p telomere and Sumatran PPY 2q (indicated by a star in Figures 4.3A and 4.4).

BACs RP11-705F7 (Figure 4.4A), RP11-642N14 (Figure 4.4B), and RP11-241K7 (Figure 4.4C) appeared to lie distal to the evolutionary breakpoint, whereas CTD-2026G6 (Figure 4.4E), RP11-413E6, RP11-441E2, RP11-73I16 (data not shown), RP11-58L10 (Figure 4.4F), RP11-214P4 (Figure 4.4G), and RP11-550F7 (Figure 4.4H) were proximal to the breakpoint. However, the abundance of duplicated DNA elements in the HSA 3p12.3 region rendered breakpoint mapping by FISH alone difficult. Only

the most distal and the most proximal BAC in the 3p12.3 contig (Figure 4.3B) produced unique hybridization signals in the orangutan and human genomes. All other clones which are closer to the 3p12.3 breakpoint hybridized to multiple paralogous sites, including HSA 3q21 and the syntenic region on PPY 2 (indicated by a white triangle in figure 4.4). The breakpoint is assumed to be located in the ~360 kb interval between RP11-642N14 and CTD-2026G6.

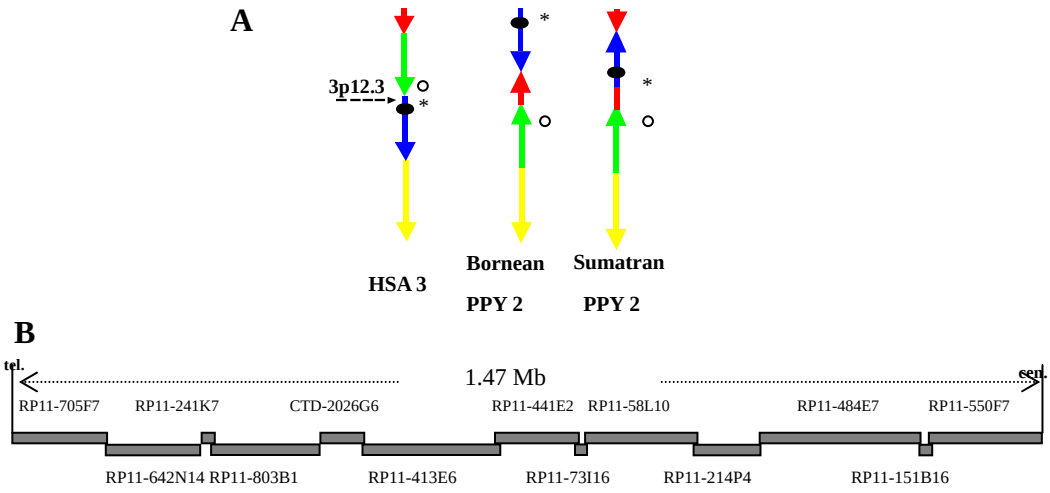



Figure 4.3. (A) Conservation of four syntenic segments (indicated by different colors) between HSA 3 and Bornean PPY 2. The dot and star symbols at the right indicate the location of the distal and proximal 3p12.3 breakpoint regions on the given chromosome. (B) Each line in the BAC contig of the HSA 3p12.3 breakpoint region represents an individual clone. The BAC name is written at the top or at the bottom.

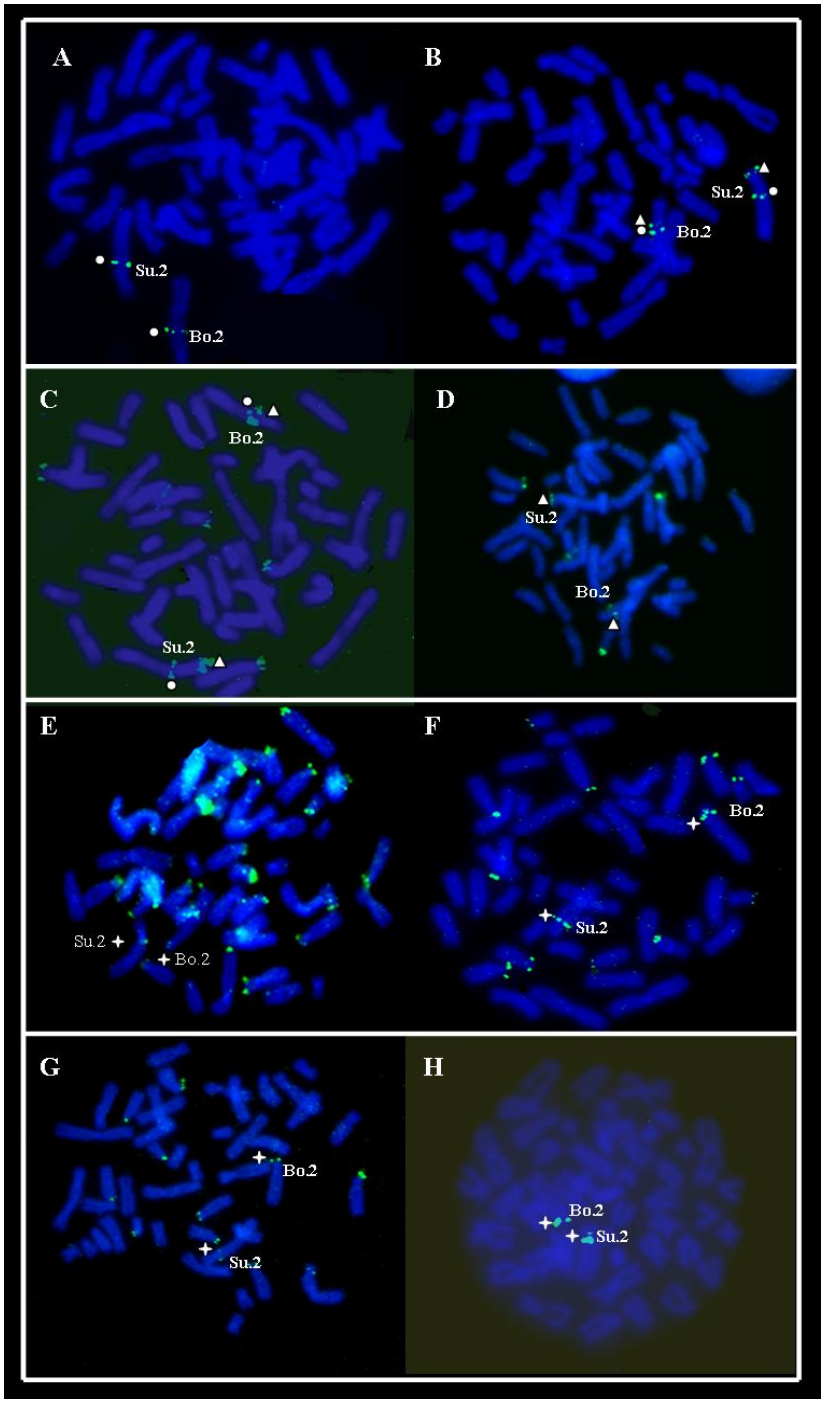



Figure 4.4. FISH with BAC clones of the 3p12.3 contig on orangutan hybrid metaphase spreads. The distal side and the proximal side of the 3p12.3 breakpoint region on PPY 2 are marked by a white dot (○) and a star (✶), respectively. The HSA 3q21-homologous region is indicated by a triangle (△). (A) BAC RP11-705F7 produced a single-copy hybridization signal distal of the 3p12.3 breakpoint. (B) BAC RP11-642N14 produced two signals on PPY 2 at the distal breakpoint and the 3p21-homologous region. (C) BAC RP11-241K7 was duplicated on PPY 2 and at multiple subtelomeric sites. (D) BAC RP11-803B1 highlighted only the 3q21-homologous site on PPY 2, but also multiple subtelomeric sites in other chromosomes. BACs CTD-2026G6 (E), RP11-58L10 (F), and RP11-214P4 (G) all hybridized to the proximal 3p12.3 breakpoint region and to multiple subtelomeric sites. (H) BAC RP11-550F7 produces a single-copy signal at the proximal breakpoint region on PPY 2.

## 4.1.3. Evolutionary breakpoint between HSA 3q21.3 and PPY 2

The evolutionary breakpoint between the blue and yellow segments (Figure 4.5A) which distinguishes HSA 3q21.3 and PPY 2 was located in the interval between YACs 896g9 and 914h1 at 130-132 Mb in HSA 3q21.3. Again a contig consisting of BACs RP11-93K22, RP11-77P16 and RP11-687B8 according to the database was assembled (Figure 4.5B). These BACs were mapped by FISH to human as well as Bornean x Sumatran orangutan hybrid chromosomes.

BAC RP11-93K22 hybridized to the proximal long arm of Bornean PPY 2 and the short arm of Sumatran PPY 2 (Figure 4.6A). This site corresponds to the tip of the blue arrow in figure 4.5A and is proximal of the evolutionary breakpoint. RP11-687B8 hybridized to the distal long arm of both Bornean and Sumatran PPY 2 (Figure 4.6C) corresponding to the end of the yellow arrow (Figure. 4.5A), distal to the breakpoint. RP11-77P16, which lies between the two breakpoint-flanking BACs in the 3q21.3 breakpoint contig (Figure 4.5B), did not produce any detectable hybridization signal on chromosome 2 in the orangutan genome, but strongly labeled the subtelomeric regions of PPY 10 and PPY 18 (Figure 4.6B). This indicates that the evolutionary rearrangement between HSA 3q21.3 and PPY 2 was not a simple breakage and reunion event. The breakpoint was located in the 317 kb interval between RP11-93K22 and RP11-687B8.

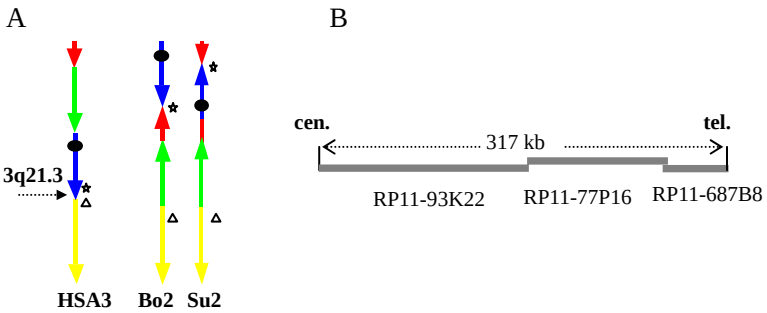



Figure 4.5. (A) Conservation of four syntenic segments (indicated by different colors) betwee HSA 3 and Bornean PPY 2. The proximal side of the 3q21.3 breakpoint region is indicated by a star (✶).The distal side of the 3q21.3 breakpoint region on PPY 2 is marked by a triangle (△). (B) BAC contig of the HSA 3q21.3 breakpoint region: Each line represents a BAC clone with “finished” sequence. BAC names are written below each line.

.

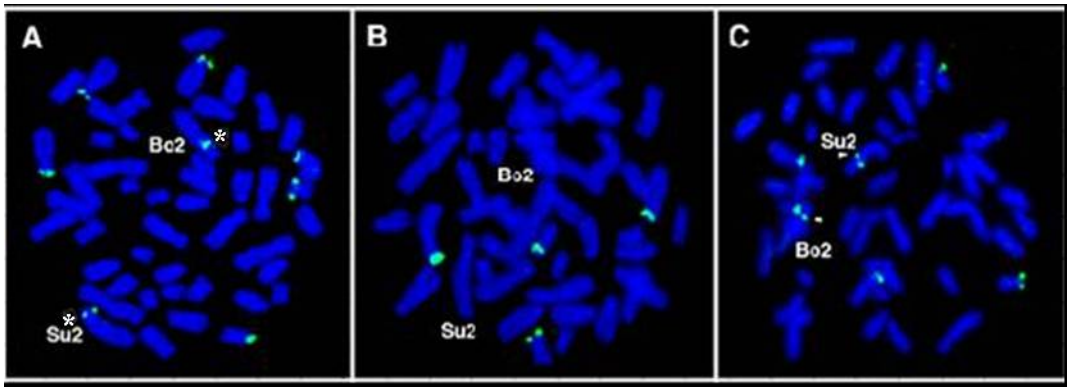



Figure 4.6. FISH with BACs RP11-93K22 (A), RP11-77P16 (B) and RP11-687B8 (C) on orangutan chromosomes. Bo2 and Su2 indicate the Bornean and Sumatran orangutan chromosome forms. The proximal side of the 3q21.3 breakpoint region is indicated by a star (✶). The distal side of the 3q21.3 breakpoint region on PPY 2 is marked by a triangle (△). (A) RP11-93K22 highlighted the proximal site to the breakpoint on PPY 2. (B) RP11-77P16 did not map to the HSA 3q21.3-syntenic site on PPY 2. (C) RP11-687B8 was located at distal to the breakpoint on PPY 2.

## 4.2. Genomic DNA insertions and deletions between HSA 3 and PPY 2

### 4.2.1. Deletion of HSA 3p12.3-syntenic sequences in the breakpoint region on PPY 2

By FISH, BAC RP11-803B1 of the breakpoint contig (Figure 4.3B) localized to both the 3p12.3- and 3q21.3-syntenic regions in human and chimpanzee (data shown in appendix - Supplementary figure 1A and B). This is consistent with an intrachromosomal duplication in a common ancestor of humans and great apes. However, in PPY 2 it did not map to the HSA 3p12.3-syntenic site. Instead, it produced a single hybridization signal on Bornean PPY 2q and Sumatran PPY 2p, which correspond to the tip of the blue arrow at HSA 3q21.3 (Figure 4.4D). I propose that a loss of RP11-803B1 sequences occurred at the HSA 3p12.3-homologous site in the orangutan lineage.

In order to confirm the FISH result and analyze the distribution of RP11-803B1 paralogous sequences in the orangutan genome, PCR typing was performed on chromosome-specific DNA libraries. Thirteen primer pairs were designed to amplify repeat-poor (supposedly unique) fragments along the entire length of BAC RP11-803B1 (Table 2.4 and Figure 4.7). Each primer pair yielded a specific amplification product from human genomic DNA. 11 of 13 primers amplified the expected fragment from a DOP-PCR library of HSA 3. Since DOP-PCR products from

flow sorted chromosomes do not represent the entire genomic sequence of this chromosome, it is not unexpected that not all 13 target sequences could be amplified. The efficiency of this PCR mapping approach appears to be in a range between 85% (an amplification product was generated from the HSA 3 library) and 95% (an amplification product was generated from at least one of the 24 human libraries tested). When the same primer pairs were used on PPY chromosome-specific DNA libraries, only two primer pairs amplified a product from PPY 2, whereas 10 gave a product from PPY 18 (homologous to HSA 16), 7 from PPY 22 (HSA 21), and 5 from PPY 10+6 (HSA 7+8) (Table 4.1). Five additional orangutan chromosomes were positive for one or two primer pairs. This implies that most RP11-803B1-homologous sequences in the orangutan genome do not reside on PPY 2.

BAC RP11-642N14 hybridized to both the 3p12.3- and 3q21-homologous sites in HSA 3 (data shown in appendix - Supplementary figure 1C), PPY 2 (Figure 4.4B), and *P. cristata* chromosome 1 (Figure 4.8E), which most closely resembles the ancestral simian chromosome (Tsend-Ayush *et al.*, 2004). This demonstrates the ancestral nature of this intrachromosomal duplication. Because BAC RP11-642N14 lies very close to the evolutionary 3p12.3 breakpoint and the unexpected hybridization results of BAC RP11-803B1, four long-range PCR products were generated from RP11-642N14 (Table 2.5 and Figure 4.7) for further FISH mapping. Similar to BAC RP11-803B1, fragments 642N14A, B and C did not hybridize to the HSA 3p12.3-homologous site on PPY 2, but were specific for the HSA 3q21-homologous site (Figure 4.8A-C). The most distal fragment 642N14D hybridized specifically to the HSA 3p12.3-homologous region on Sumatran and Bornean PPY 2q (Figure 4.8D), corresponding to the green arrowhead in figure 4.3A. Evidently, orangutan sequences homologous to the proximal part of RP11-642N14 do not reside at the HSA3p12.3-homologous site in the orangutan genome.

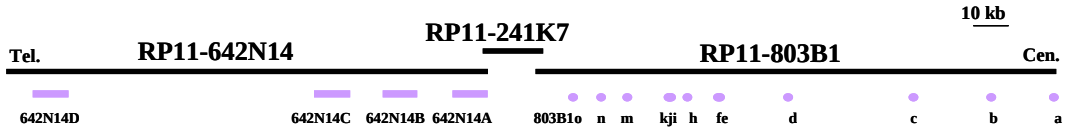
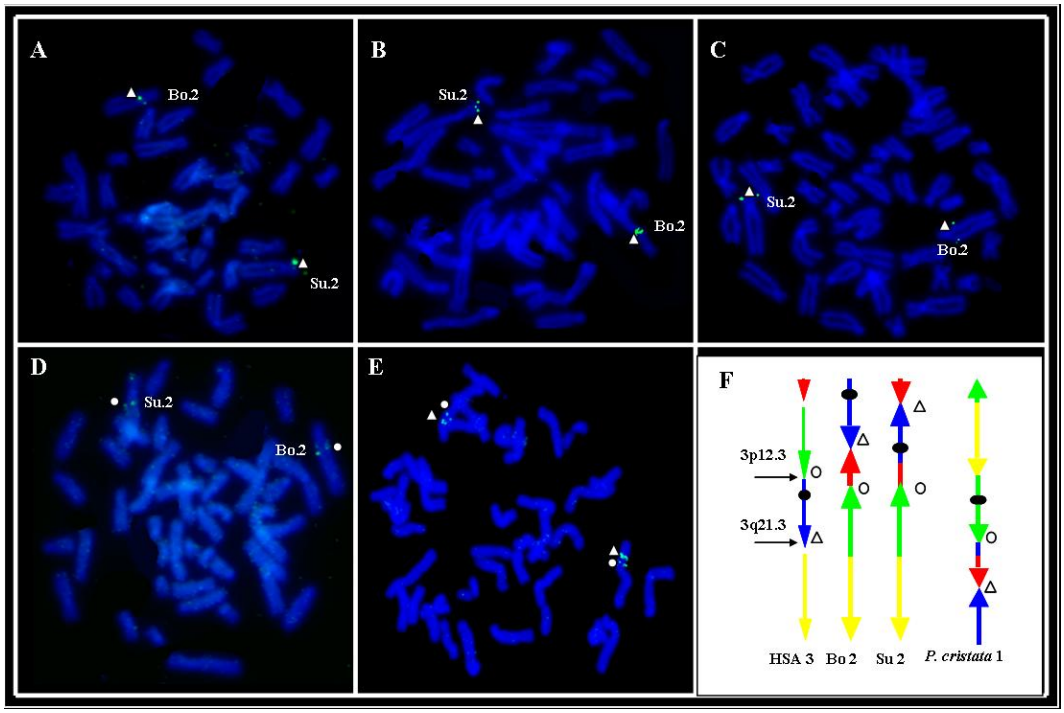



Figure 4.7. Genomic position of PCR fragments from RP11-803B1 (purple circles) and RP11-641N14 (purple bars) used for PCR analysis of PPY chromosome libraries and FISH, respectively, in order to narrow down the 3p12.3 breakpoint region. The primer sequences are listed in tables 2.4 and 2.5.

Table 4.1. PCR analysis of PPY chromosome-specific DNA libraries. PCR fragments are listed from proximal to distal.

| PCR fragment | PPY chromosome library (homologous human chromosome) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 2 (III) | 3 (IV) | 10+6 (VII +VIII) | 13 (IX) | 16 (XV) | 17 (XVIII) | 18 (XVI) | 21 (XX) | 22 (XXI) |
| 803B1a | | | + | | | | + | | |
| 803B1b | | | + | | | | + | + | |
| 803B1c | | | | | | + | + | | |
| 803B1d | | | | | | | + | | |
| 803B1e | | | | | | | + | | + |
| 803B1f | | | | | | | + | | + |
| 803B1h | | | | | | | | | + |
| 803B1i | | | | | | | + | | |
| 803B1j | | | + | | | | + | | |
| 803B1k | + | + | | | + | | + | | + |
| 803B1m | | | | | | | + | | + |
| 803B1n | | | + | + | | | + | + | + |
| 803B1o | + | | + | + | | | | | + |

Only PPY chromosomes that are positive with at least one primer pair are listed.




Figue 4.8. (A-D) FISH with long-range PCR fragments of BAC RP11-642N14 (see table 4.2) on Bornean-Sumatran orangutan hybrid metaphase spreads. The position corresponding to the distal side of the 3p12.3 breakpoint (tip of the green bar in F) is marked by a white dot (○), the HSA 3q21-homologous region (tip of the blue bar) by a triangle (△). Fragments 642N14A (A), 642N14B (B) and 642N14C (C) hybridized to the HSA 3q21-homologous region on Sumatran (Su) PPY 2p and Bornean (Bo) PPY 2q, whereas fragment 642N14D (D) hybridized to the distal 3p12.3 breakpoint region of Sumatran and Bornean PPY 2q. (E) BAC RP11-642N14 highlighted the HSA 3p12.3 and 3q21-homologous sites on *P. cristata* chromosome 1. (F) Conservation of four syntenic segments (indicated by different colors) among HSA 3, Bornean PPY 2 and *P. cristata* chromosome 1.

## 4.2.2. Insertion of a DNA segment at the 3q21 breakpoint in the human and African ape lineage

BAC RP11-77P16 hybridized to syntenic sites on HSA 3q21.3 and PTR 2 (data shown in appendix - Supplementary figure 1E and F), but did not FISH map to the homologous orangutan (Figure 4.6B), gibbon, and silvered-leaf monkey (data not shown) chromosomes. PCR typing of human and orangutan chromosome specific libraries was performed with 31 PCR primer pairs which amplify repeat-poor (supposedly unique) 100-300 bp fragments along the entire length of the breakpoint region (Table 2.6 and Figure 4.11A). Seven of the 31 primer pairs tested did not work with orangutan genomic DNA. This could be due to microdeletions, which occurred relatively frequently during primate genome evolution (Frazer *et al.*, 2003), or to the sequence divergence between humans and orangutan. DNA fragments from proximal

RP11-93K22, distal RP11-77P16, and all of RP11-687B8 were amplified from PPY 2, whereas fragments from distal RP11-93K22 and a large part of RP11-77P16 were mainly amplified from other PPY chromosomes which are not homologous to HSA 3 (Table 4.2 and Figure 4.11C). This assay demonstrated that only 6-7% of RP11-77P16-homologous sequences reside in the 3q21.3-breakpoint region on PPY 2. This strongly suggest that an insertion of an approximately 200 kb duplicated DNA segment (corresponding to a large part of RP11-93K22 and almost entire RP11-77P16) has occurred during the 3q21 rearrangement that led to the human and chimpanzee chromosome form, after divergence of the orangutan lineage from humans and African apes 12-16 million years ago. This PCR mapping also allowed me to map the breakpoint region between HSA 3q21.3 and PPY 2 within a 223 kb interval between PCR fragments 93K22q and 77P16i (Figure 4.11C). Consistent with these FISH-mapping results, at least 211 kb of the breakpoint region (from 93K22l to 77P16h) containing 108 kb of RP11-93K22 and 103 kb of RP11-77P16 sequences were duplicated on PPY 10 and 18, but not present on PPY 2. Thus, the HSA 3q21.3-syntenic breakpoint must be localized in the 4.6 kb PPY interval 1 between 93K22q and 93K22l or in the 7.6 kb PPY interval 2 between 77P16h and 77p16i.

## 4.3. Hot spots for chromosome evolution

### 4.3.1. Reuse of the HSA 3q21.3-syntenic breakpoint region in independent evolutionary chromosome rearrangements

In the siamang gibbon, a breakpoint between YACs 896g9 and 754c8 at 130-134 Mb in HSA 3q21.3 led the formation of the HSA 3 homologous parts on siamang gibbon (*Hylobates syndactylus*, HSY) chromosomes 10 and 21 (Figures 1.4 and 4.9). In Old World monkeys, the inversion between the simian HSA 3 ancestor and silvered leaf monkey (*P. cristata*) chromosome 1 involved a breakpoint between YACs 904d6 and 914h1 at 127.3-132 Mb in HSA 3q21.3 (Figures 1.4 and 4.9). In order to further narrow down these two independent HSA 3q21-syntenic breakpoints, BACs RP11-93K22, RP11-77P16 and RP11-687B8 were FISH mapped on siamang gibbon and *P. cristata* chromosomes. In siamang, RP11-93K22 and RP11-687B8 hybridized to HSY 21 and HSY 10 (Figure 4.9), respectively, flanking a HSY-specific translocation breakpoint. In Old World monkeys, RP11-93K22 and RP11-687B8 mapped proximal and distal of an inversion breakpoint leading to *P. cristata* chromosome 1 (Figure 4.9). Similar to the situation in orangutan (see chapters 4.1.3 and 4.2.2), RP11-77P16 did not FISH-map to

a siamang or silvered-leaf monkey chromosome that shares conserved synteny with HSA 3q21.3 (data not shown). Thus the breakpoint region between RP11-93K22 and RP11-687B8 was not only used in the orangutan lineage, but also involved in two independent chromosomal rearrangements in the gibbon and Old World monkey lineages.

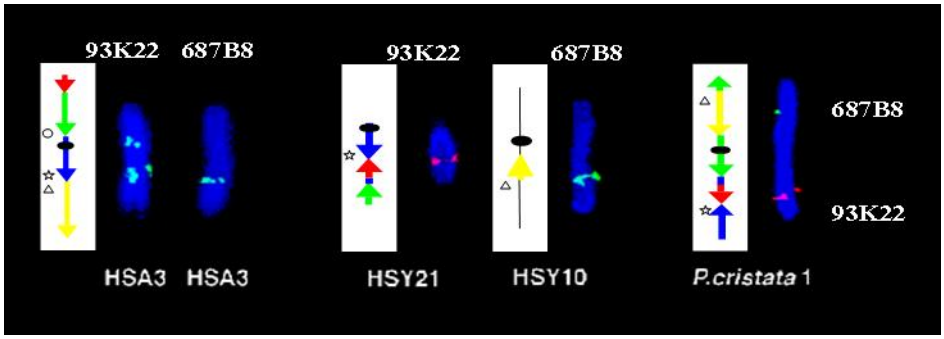



Figure 4.9. FISH with BAC clones RP11-93K22 and RP11-687B8 on HSA 3, HSY 10 and 21, and *P. cristata* 1. In contrast to RP11-687B8 which is specific for HSA 3q21.3, RP11-93K22 also highlighted the paralogous site on HSA 3p12.3 (indicated by a dot). In the siamang gibbon, the two clones mapped to HSY 21 and HSY 10 (indicated by a star and a triangle, respectively), flanking a siamang-specific translocation breakpoint. In Old World monkey, RP11-93K22 and RP11-687B8 hybridized to different arms of *P. cristata* 1 and, thus, flank an independent HSA 3q21.3-syntenic inversion breakpoint.

To confirm and refine these FISH results, PCR typing of siamang chromosome specific libraries was performed with the same primer pairs examined on orangutan chromosomes (Table 2.6). In the siamang genome, only two primer pairs did not work. Most DNA fragments from proximal RP11-93K22 were amplified exclusively from HSY 21, whereas fragments from distal RP11-77P16 and RP11-687B8 were specific for HSY 10 (Table 4.2). Thus, the HSA 3q21.3-syntenic translocation breakpoint lies in the 220 kb interval between PCR fragments 93K22v and 77P16j (Figure 4.11D). At least 108 kb of RP11-93K22 and 92 kb of RP11-77P16 sequences from the breakpoint region (from 93K22l to 77P16p) were PCR mapped to HSY 1 and HSY 13/14, but not to a HSA 3q21.3-homologous siamang chromosome. This narrowed the breakpoint to the 7.5 kb HSY interval 1 between 93K22v and 93K22l or to the 12.7 kb HSY interval 2 between 77P16p and 77P16j.

Table 4.2. PCR analysis result of BAC fragments from the 3q21.3 breakpoint contig (table 2.6) on human, orangutan, and siamang chromsome-specific labraries.

| BAC frag-ment | Human genome | | Orangutan genome | | Siamang genome | |
|---|---|---|---|---|---|---|
| | Genomic DNA | Chromosome library | Genomic DNA | Chromosome library | Genomic DNA | Chromosome library |
| 93K22d | + | - | + | - | + | HSY 21 |
| 93K22s | + | HSA 3, 13 | + | PPY 2, 20, 21 | + | N.d. |
| 93K22c | + | HSA 3 | - | N.d. | + | HSY 21 |
| 93K22e | + | - | - | N.d. | + | HSY 21 |
| 93K22g | + | HSA 3 | - | N.d. | + | - |
| 93K22h | + | - | - | N.d. | + | HSY 21 |
| 93K22r | + | HSA 3 | + | PPY 2 | + | HSY 12, 21 |
| 93K22v | + | HSA 3 | + | PPY 2 | + | HSY 21 |
| 93K22q | + | HSA 3 | + | PPY 2 | - | N.d. |
| 93K22p | + | HSA 9, 16 | + | PPY 21, 22 | + | - |
| 93K22l | + | HSA 3, 4, 7, 8, 11 | + | PPY 1, 10, 14, 18 | + | HSY 1, 13/14 and more |
| 93K22k | + | HSA 3, 4, 7, 10, 11 | - | N.d. | + | HSY 1, 13/14 |
| 93K22j | + | HSA 3, 4, 7, 8, 11 and more | + | PPY 2, 10, 18 and more | + | HSY 1, 13/14 and more |
| 93K22i | + | HSA 3, 4, 7, 11 and more | + | PPY 2, 10, 13, 17, 18 | + | HSY 1, 13/14 and more |
| 93K22b | + | HSA 4, 7, 11, 21 | + | PPY 10, 18 | + | HSY 1, 3/5 |
| 93K22a | + | HSA 3 | + | PPY 10, 18 | + | HSY 1, 22/24 |
| 77P16a | + | HSA 3, 4, 8, 10, 11, 16 | + | PPY 2, 5, 10, 11, 13, 18 | + | HSY 1, 13/14 |
| 77P16b | + | HSA 3, 4, 7, 8, 11, 16 and more | + | PPY 10, 18 | + | HSY 1, 3/5 |
| 77P16d | + | HSA 3, 4, 7, 8, 11, 12, 15, 16 | + | PPY 10, 18 | + | HSY 1 and more |
| 77P16f | + | HSA 4, 7, 8, 11 | + | PPY 10, 18 | + | HSY 1, 13/14 |
| 77P16g | + | HSA 3, 4, 11 | + | PPY 5, 10, 13, 18 | + | HSY 1, 13/14 |
| 77P16p | + | HSA 1, 2. 3, 8, 11, 16 | + | - | + | HSY 1 |
| 77P16h | + | HSA 3, 4, 8, 11, 19 | + | PPY 10, 18 | + | - |
| 77P16j | + | HSA 3 | + | PPY 13 | + | HSY 10 |
| 77P16k | + | HSA 3 | + | PPY 1, 2, 4, 5, 13, 17, 21 | + | HSY 10 |
| 77P16i | + | HSA 3, 13 | + | PPY 2 | + | HSY 10 |
| 687B8d | + | HSA 3, 15, Y | + | PPY 2, 16 | + | HSY 10, 13/14, 16/19 |
| 687B8e | + | HSA 3 | - | N.d. | - | N.d. |
| 687B8g | + | HSA 3 | + | PPY 2 | + | HSY 10 |
| 687B8b | + | - | - | N.d. | + | HSY 10 |
| 687B8c | + | HSA 3 | + | PPY 2 | + | HSY 10 and more |

N.d., not done

## 4.3.2. Reuse of the evolutionary breakpoint regions between human chromosome 3 and orangutan 2 during vertebrate evolution

Comparison of the human genome sequence with current assemblies of the mouse, rat, and chicken genomes revealed that rearrangements of the HSA 3q21.3-syntenic region also occurred in rodents and birds. In the mouse (*Mus musculus*, MMU) genome, the 3q21.3 breakpoint cluster region was involved in a rearrangement between MMU 6 and MMU 9, in the rat (*Rattus norvegicus*, RNO) it diverged on RNO 4 and RNO 8, and in the chicken (*Gallus gallus*, GGA) it was distributed to GGA 12 and 9 (Figure 4.10). Similarly, the two other inversion breakpoints between PPY 2 and HSA 3, which were located between BACs RP11-642N14 and CTD-2026G6 at HSA 3p12.3 and to BAC RP11-616M11 at HSA 3p25.1, were also in regions of breaks of synteny (BOS) between the human, rodent, and/or avian genomes (Figure 4.10). This indicates that the three breakpoint contigs on human chromosome 3 represent hot spots for break of synteny between vertebrate genomes.

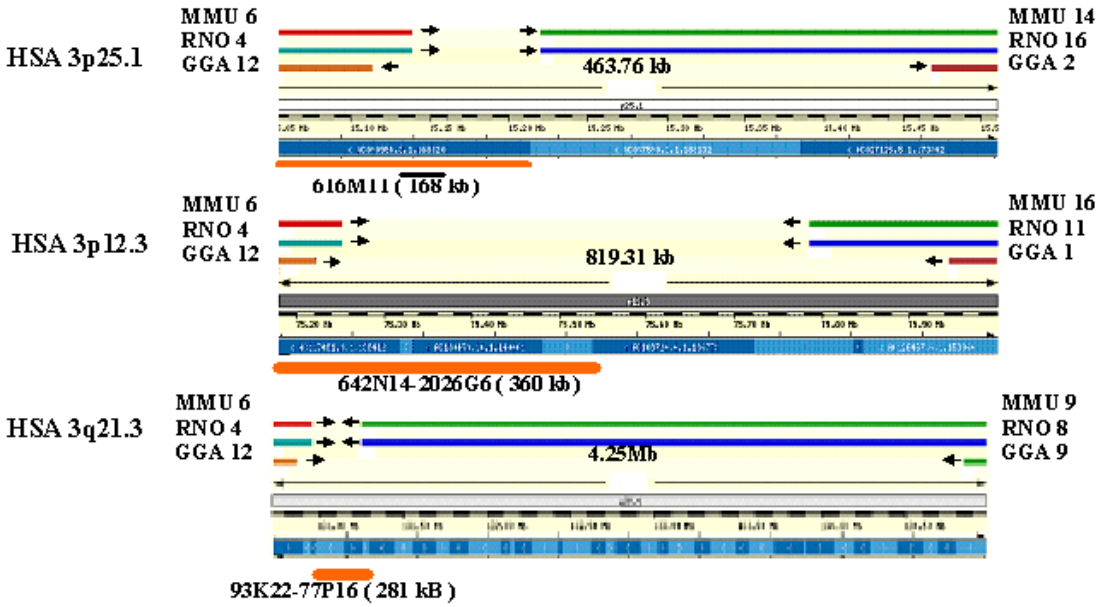



Figure 4.10. The three evolutionary breakpoints between HSA 3p25.1, 3p12.3, and 3q21.3 and PPY 2 correspond to breaks of synteny between the human genome and the mouse, rat, and chicken genomes. Homology maps were delineated by the Wellcome Trust Sanger Institute ensembl genome browser. The 3p25.1 breakpoint in BAC 616M11 (indicated by black line below the BAC line) breaks the synteny between mouse 6 and 14, rat 4 and 16, and chicken 12 and 2. The 3p12.3 breakpoint lies in the paralogous contig between BACs RP11-642N14 and CTD-2026G6 and corresponds to a BOS between mouse 6 and 16, rat 4 and 11, and chicken 12 and 1. The 3q21.3 breakpoint cluster in BACs RP11-93K22 and RP11-77P16 corresponds to a BOS between mouse 6 and 9, rat 4 and 8, and chicken 12 and 9.

# 4.4. Evolutionary breakpoints are associated with segmental duplications and repetitive elements

## 4.4.1 Paralogous segments at the HSA 3q21.3 and 3p12.3 breakpoint regions

BACs RP11-93K22 and RP11-77P16 from HSA 3q21.3 also hybridized to 3p12.3 and to multiple subtelomeric and pericentromeric sites in the human genome (data shown in appendix – Supplementary figure 1D and 1E). PCR mapping was used to analyze the distribution of 3q21.3-paralogous segments in human genome. When tested on degenerate oligonucleotide primed (DOP) PCR products of flow-sorted human chromosomes, four out of 31 primer pairs did not give any product, 11 pairs amplified sequences exclusively from HSA 3, 13 pairs amplified sequences from both HSA 3 and multiple other chromosomes, and 3 pairs amplified sequences from a chromosome(s) which was not HSA 3 (Tables 2.6 and 4.2). The efficiency of PCR mapping was in the range between 77% (an amplification product was generated from the HSA 3 library) and 87% (an amplification product was generated from at least one of the 24 human libraries tested). The results of PCR typing demonstrate that only the most distal and the most proximal fragments of the 3q21.3 breakpoint contig are specific for HSA 3, whereas sequences closer to the breakpoint were duplicated at multiple other chromosomes, in particular on HSA 4, 7, 8 and 11 (Figure 4.11B and Table 4.2).

Computational sequence analysis revealed several RP11-93K22 and RP11-77P16 paralogous segments in the human genome (Table 4.3), including two paralogous blocks on HSA 3. Intriguingly the 3p12.3 breakpoint contig comprises >180 kb of DNA with 90% sequence similarity to the 3q21.3 breakpoint cluster. Contig 3q21.2 consists of BACs RP11-379B18 and RP11-666A20 and lies approximately 4 Mb proximal of the 3q21.3 breakpoint cluster. It contains >110 kb DNA with 90% sequence similarity to the breakpoint cluster. Nine additional paralogous blocks are located in the subtelomeric regions of HSA 4, 7, 8, 11, 12 and 16 (Table 4.3). This result is consistent with the FISH and PCR mapping results. Except the paralogous blocks on chromosome 16, all other paralogous segments including the 3q21.3 breakpoint cluster contain olfactory receptor genes of the 7E subfamily, which may predispose to segmental duplications (Newman and Trask, 2003).

Table 4.3. DNA segments (BAC contigs) paralogous to the 3q21.3 breakpoint contig

| 3p12.3 (75.3-75.7 Mb) | 3q21.2 (126.8-127 Mb) | 4p16.3 (4-4.4Mb) | 4p16.1 (9.2-9.5 Mb) | 7p22.1 (6.6-6.8 Mb) | 8p23.1-A (7.87-8.2 Mb) | 8p23.1-B (12.2-12.6 Mb) | 11q13.2 (67.1-67.6 Mb) | 11q13.4 (70.9-71.3 Mb) | 12p13.3 (8.25-8.52 Mb) | 16p13.3 (5.06-5.22 Mb) |
|---|---|---|---|---|---|---|---|---|---|---|
| RP11-803B1[a] 134 kb 90.7% | CTA-388B5 38.5 kb 91% | RP11-489M13 104 kb 90.8% | RP11-264E23 105 kb 91% | RP11-740N7 108 kb 91% | RP11-52B19 59.2 kb 90% | RP11-419I17 104 kb 92% | RP11-655M14 54 kb 91% | RP11-684B2 47.3 kb 91% | RP11-90D4 116 kb 93.6% | RP11-10K17 52 kb 90% |
| CTD-2026G6 17.5 kb 89.6% | RP11-379B18 88.4 kb 90.1% | RP11-324I10 22.3 kb 91% | RP11-180A12 56.1 kb 90% | | RP11-556O5 90.8 kb 92.1% | RP11-303G3 36 kb 90% | RP11-119D9 131 kb 92.5% | RP11-167J8 86.9 kb 93.6% | RP11-113C12 45 kb 90% | |
| RP11-413E6 32 kb 90.8% | RP11-666A20 28.1 kb 90.6% | | | | | | RP5-901A4 32 kb 90% | CTD-2313N18 38.5 kb 91% | | |

[a] The numbers below the BAC names indicate the length of the paralogous DNA segment and the percentage of sequence similarity.

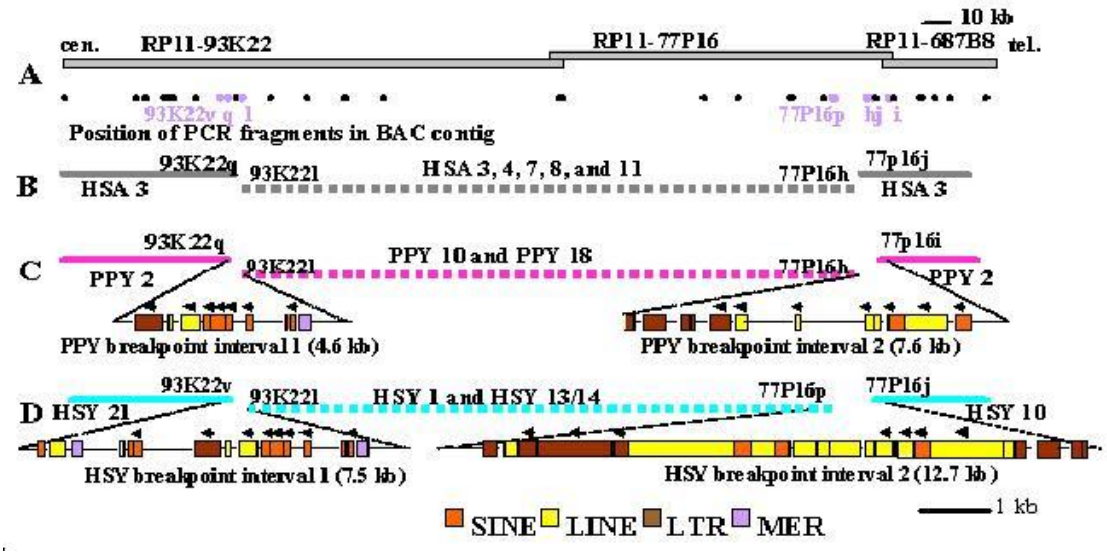



Figure 4.11. PCR analysis of the HSA 3q21.3-syntenic breakpoint region. (A) Location of PCR fragments (indicated by filled circles) which were tested for their presence or absence in a given DNA library (Tables 4.3 and 4.4) along the 3q21.3 breakpoint contig. The PCR fragments 93K22v, q,and l and 77P16p, h, j and i, which delineate the breakpoint intervals in orangutan (C) and siamang gibbon (D), are indicated in purple. (B) PCR typing of flow-sorted human chromosomes revealed that the DNA sequences between fragments 93K22l and 77p16h are duplicated on HSA 3, 4, 7, 8 and 11 (gray dotted line). (C) The orangutan-specific inversion breakpoint between HSA 3q21.3 and PPY 2 is located in the 4.6 kb PPY breakpoint interval 1 between PCR fragments 93K22q and 93K22l or in the 7.6 kb PPY interval 2 between 77P16h and 77p16i. The differently colored bars in the enlarged breakpoint intervals indicate the presence of repetitive SINE, LINE, LTR and MER sequences. The 211 kb segment between 93K22l and 77P16h (pink dotted line) is not present on the HSA 3-syntenic PPY 2, but has been duplicated on PPY 10 and 18. (D) The HSA 3q21.3-syntenic translocation breakpoint leading to HSY 21 and HSY 10 lies in the 7.5 kb HSY breakpoint interval 1 between 93K22v and 93K22l or in the 12.7 kb interval 2 between 77P16p and 77P16j. Similar to the situation in orangutan, sequences between 93K22l and 77P16g were not found on the derivative translocation chromosomes, but have been duplicated on HSY 1, 13 and/or 14 (light blue dotted line).

## 4.4.2 Paralogous segments of the HSA 3p25.1 breakpoint contig

In addition to the split signals on PPY 2, the two breakpoint spanning BACs RP11-616M11 and RP11-72K12 also highlighted subtelomeric regions of PPY 10 and PPY18 (Figure 4.2C and 4.2D). These two clones also hybridized to multiple sites in the human genome (data shown in appendix - Supplementary figure 1G and H). The 3p25.1-paralogous segments were studied *in silico* using the Human Genome Segmental Duplication Database (http://projects.tcag.ca/humandup. Biocomputational analysis revealed that a ~21 kb segment (corresponding to basepairs 50863-72000 of RP11-616M11) within the breakpoint region has been duplicated to at least 15 sites in the human genome (Table 4.4). Segmental duplications which are shared among the three evolutionary breakpoints between HSA 3 and PPY 2 are shown in Figure 4.12.

Table 4.4. Paralogous segments of the HSA 3p25.1 breakpoint region in the human genome (May 2004 assembly).

| Cytogenetic band | Clone | Chromosome position | Percent identity | Alignment length (bp) |
|---|---|---|---|---|
| 3p12.3 | CTD-2026G6 | chr3: 75669682… 75683137 | 95.48% | 13398 |
| | | chr3: 75691841… 75698385 | 95.13% | 5420 |
| 3q21.2 | RP11-379B18 | chr3: 126981587… 127001689 | 95.85% | 20103 |
| 4p16.1 | RP11-264E23 | chr4: 9223662… 9246110 | 94.71% | 21138 |
| 4p16.3 | RP11-489M13 | chr4: 4194211… 4214008 | 95.24% | 19798 |
| 7q21.3 | RP11-380G21 | chr7: 97171923… 97189809 | 95.95% | 17255 |
| 8p23.1 | RP11-158L15 | chr8: 6950663… 6972817 | 96.17% | 20850 |
| | | chr8:7031431… 7053225 | 94.6% | 21138 |
| 8p23.1 | RP11-52B19 | chr8: 7974160… 7995935 | 94.43% | 21138 |
| 8p23.1 | RP11-303G3 | chr8: 12467982… 12489696 | 94.55% | 21138 |
| | | chr8:12526100… 12547603 | 95.06% | 20830 |
| 11p15.4 | CTD-2502D10 | chr11: 3523280… 3545688 | 96.08% | 21138 |
| 11q13.2 | RP11-138N3 | chr11: 67445429… 67466263 | 96.74% | 20835 |
| 11q13.4 | RP11-167J8 | chr: 71041540… 71063260 | 96.32% | 21138 |
| 12p13.31 | RP11-113C12 | chr12: 8399536… 8420819 | 95.24% | 19948 |

### 4.4.3. Evolutionary breakpoints are associated with repetitive elements

The frequency and classes of repeated elements in evolutionary breakpoint regions was analyzed with RepeatMasker (http://menu.hgmp.mrc.ac.uk/menu-bin/Nix/Nix.pl). The 3p25.1 breakpoint region consists of 39% repeats. The breakpoint contigs at HSA 3p12.3 and HSA 3q21.3 have an even higher percentage of repetitive elements: 56.3% and 63.0% of sequences are marked by RepeatMasker, respectively. Both the HSA 3p12.3 and the 3q21.3 contig contain approximately 2% satellite DNAs. However, the repeat composition of the three delineated breakpoint-containing intervals varies considerably (Table 4.5). Generally SINEs and/or LTRs are enriched compared with the average human genome (The International Human Genome Sequencing Consortium 2001). In the HSY breakpoint interval 2 the percentage of LINE elements is much higher than that of other elements and the average human genome.

Table 4.5. Repetitive elements of evolutionary breakpoint regions. The shaded numbers indicate a higher percentage compared with the average human genome.

| | | | SINE | LINE | LTR | DNA transposons |
|---|---|---|---|---|---|---|
| 3p25.1 breakpoint region (28.7 kb) (RP11-616M11 and RP11-72K12) | | | **20.3%** | **7.8%** | **9.6%** | **1.6%** |
| 3p12.3 breakpoint region (RP11-642N14 to CTD-2026G6; 359 kb) | | | **15.5%** | **15.0%** | **21.3%** | **1.5%** |
| 3q21.3 | breakpoint cluster region (226 kb) | 93K22v-77P16i | **18.5%** | **12.4%** | **26.7%** | **2.6%** |
| | PPY breakpoint interval 1 (4.6 kb) | 93K22q-93K22l | **20.6%** | **11.9%** | **11.8%** | **5.0%** |
| | PPY breakpoint interval 2 (7.4 kb) | 77P16h-77P16i | **8.0%** | **19.5%** | **15.6%** | **0.0%** |
| | HSY breakpoint interval 1 (7.5 kb) | 93K22v-93K22l | **18.3%** | **11.3%** | **7.3%** | **5.4%** |
| | HSY breakpoint interval 2 (12.7 kb) | 77P16h-77P16j | **7.4%** | **49.7%** | **26.7%** | **0.0%** |
| Average human genome | | | **13.1%** | **20.4%** | **8.3%** | **2.8%** |

## 4.5. Gene contents of the three evolutionary breakpoint regions between human chromosome 3 and orangutan chromosome 2

The National Center for Biotechnology Information (NCBI) database was used to determine the gene content of the three evolutionary breakpoint regions. There are six bona fide genes and one predicted gene in the ~500 kb interval flanking the inversion breakpoint at HSA 3p25.1. Among these genes, the zinc finger, *FYVE domain containing 20* (*ZFYVE20*) gene maps ~30 kb distal to the breakpoint and the predicted gene LOC344875 (similar to RIKEN cDNA E330026B02) lies immediately adjacent but proximal of breakpoint. The *calpain 7* (*CAPN7*) gene is the most proximal (~80 kb to the breakpoint) validated gene (Figure 4.12 and Table 4.6).

The 3p12.3 breakpoint region contains eight hypothetical gene loci and four olfactory receptor pseudogenes. All hypothetical genes are supported by ESTs or ambiguously aligned mRNAs (Figure 4.12 and Table 4.6).

So far, five hypothetical genes, LOC401389, LOC442092, LOC440976, LOC440977 and LOC440978, and two olfactory receptor pseudogenes, *OR7E129p* and *OR7E21p*, have been mapped within the HSA3q21.3 breakpoint cluster region (Figure 4.12 and Table 4.6). One validated gene, *TRH*, lies immediately adjacent to, but outside the rearranged region. Interestingly, the genomic region of the hypothetical protein encoded by LOC440978 spans the distal rearrangement and duplication breakpoint (Figure 4.12).

Table 4.6. Gene contents of three evolutionary breakpoints contigs (according to the NCBI database).

| *Breakpoint* | *Gene* | *Description* | *Evidence** |
|---|---|---|---|
| 3p25.1 (from RP11-316A10 to RP11-321I9) | FGD5 | FYVE, RhoGEF and PH domain containing 5 | C |
| | NR2C2 | nuclear receptor subfamily 2, group C, member 2 | C |
| | MRPS25 | mitochondrial ribosomal protein S25 | C |
| | ZFYVE20 | zinc finger, FYVE domain containing 20 | C |
| | LOC344875 | similar to RICKEN cDNA E330026b02 | P |
| | CAPN7 | calpain 7 | C |
| | SH38P5 | SH3-domain binding protein 5 (BTK-associated) | C |
| 3p12.3 (from RP11-642N14 to RP11-413E6) | LOC440964 | LOC440964 | E |
| | OR7E66P | olfactory receptor, family 7, subfamily E, member 66 pseudogene | C |
| | OR7E22P | olfactory receptor, family 7, subfamily E, member 22 pseudogene | C |
| | OR7E55P | olfactory receptor, family 7, subfamily E, member 55 pseudogene | C |
| | LOC339879 | Similar to beta-1,4-mannosyltransferase; beta-1,4 mannosyltransferase | I |
| | LOC440965 | hypothetical gene supported by AK057279 | I |
| | OR7E121P | olfactory receptor, family 7, subfamily E, member 121 pseudogene | C |
| | LOC285296 | similar to unc-93 homolog B1; unc93 (C.elegans) homolog B; unc-93 related protein; unc93 (C.elegans) homolog B1 | I |
| | LOC391552 | similar to angiogenic factor VG5Q; vasculogenesis gene on 5q; VG5Q protein; (pseudogene) | ? |
| | LOC285299 | similar to FSHD region gene 2 protein | I |
| | LOC401074 | hypothetical LOC401074 | I |
| | LOC377064 | kruppel-like zinc finger factor X17 | C |
| 3q21.3 (from RP11-93K22 to RP11-77P16) | TRH | thyrotropin-realeasing hormone | C |
| | LOC401389 | hypothetical LOC401389 | E |
| | OR7E129P | olfactory receptor, family 7, subfamily A, member 129 pseudogene | C |
| | OR7E21P | olfactory receptor, family 7, subfamily E, member 21 pseudogene | C |
| | LOC442092 | similar to beta-1,4-mannosyltransferase; beta-1,4 mannosyltransferase | P |
| | LOC440976 | similar to hypothtical protein SB153 isoform 1 | P |
| | LOC440977 | LOC440977 | E |
| | LOC440978 | similar to alpha 3 type VI collagen isoform 2 precursor; collagen VI, alpha-3 | E |

*Evidence code: C, confirmed gene model based on alignment of mRNA or mRNA plus ESTs to the genomic sequence. E, model based on EST evidence only. P, model predicted by GenomeScan. ?, discrepancy between the mRNA sequence and the gene model. I, interim locus ID, model based on alignment of mRNAs or mRNAs plus ESTs, to the genomic sequence, in which the aligning transcripts could not be unambiguously assigned to a preexisting Locus ID.

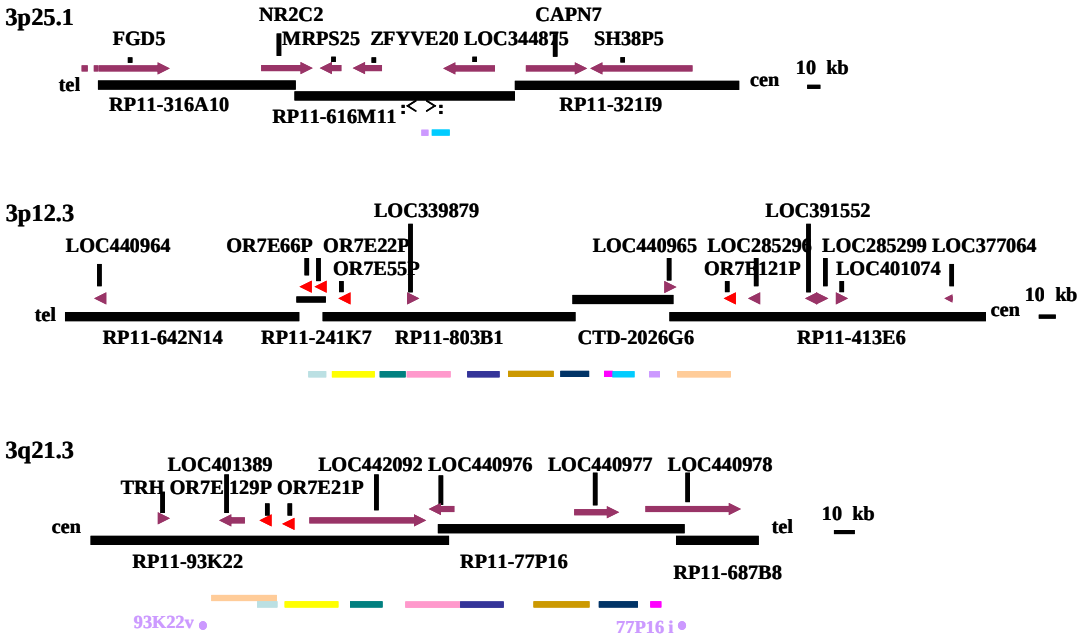



Figure 4.12. Genomic structure and duplicative nature of three evolutionary breakpoint regions on HSA 3. The breakpoint interval at 3p25.1 is indicated by two vertical dash lines below the BAC contig. The breakpoint at 3p12.3 is between RP11-642N14 and CTD-2026G6. The breakpoint at 3q21.3 is between PCR fragments 93K22v and 77P16i indicated by two purple dots below the respective BACs. Gene sequences in the breakpoint contigs are indicated by arrows in the 5'-3' direction of the sense strand above of the BAC contigs. Each gene is indicated by its NCBI name. The 5' of FGD6 is distal to RP11-316A10 in the 3p25.1 contig and indicated by a dotted line. Olfactory receptor genes are indicated by red arrowheads. Paralogous sequence blocks (>90% sequence similarity) are indicated by identical colors and forms of lines below the BAC contigs. The whole duplicated segments between 3p12.3 and 3q21.3 are >100 kb long, whereas the 3p25.1contig contains only approximately 19 kb of duplicated sequence paralogous to CTD-2026G6 at 3p12.3.

## 4.6. The *ROBO2* gene at HSA 3p12.3 and paralogous sequences

### 4.6.1. Identification of a new *ROBO2* isoform

Three BAC clones, RP11-73I16, RP11-58L10 and RP11-214P4, of the HSA 3p12.3 breakpoint region (Figure 4.3) produced multiple signals in the siamang gibbon (Figure 4.13) and orangutan genomes. The contig of these three BACs is 255 kb long and lies ~470 kb proximal to the HSA 3p12.3 evolutionary breakpoint. RP11-73I16 hybridized to human 3p12.3 and the telomeric regions of four chromosome pairs. RP11-58L10 hybridized to human 3p12.3 and weakly to the centromeric regions of two chromosome pairs. RP11-214P4 produced a specific signal on human chromosome 3 (data shown in appendix - Supplementary figure 1I-1K). Blast analysis

of the human EST database revealed one IMAGE clone, IMAGE5767984, located in RP11-58L10 and RP11-797J6, which lies 1.16 Mb proximal to RP11-58L10. Sequence analysis revealed that 388 nucleotides at the 5' end of of IMAGE5767984 have no similarity to any known human gene, whereas basepairs 389-832 are 95% identitical to the second and third exons of the *ROBO2* gene (Ensembl Gene ROBO2: ENSG00000185008). In order to study the relationship between IMAGE5767984 and *ROBO2*, RT-PCR analysis was performed on human fetal brain RNA. The reverse primer (no.1 in table 2.7) for cDNA synthesis was located in exon 10 of *ROBO2*. Two forward primers which are specific for the first exon of IMAGE5767984 (no.2 in table 2.7) and *ROBO2* (no.3 in table 2.7) (Figure 4.14A) were used for subsequent PCR. The resulting PCR products were cloned into pCR®2.1-TOPO® vector and sequenced. A specific product of 1698 bp was amplified from human fetal RNA using the IMAGE clone specific forward primer and *ROBO2* reverse primer. A 1471 bp band was amplified with the *ROBO2* forward primer. Previously, two isoforms were identified for *ROBO1*, another member of the *ROBO* gene family (Clark *et al.*, 2002). Similar to the two human *ROBO1* isoforms, the newly discovered isoform of *ROBO2* was named *ROBO2a* and the known isoform *ROBO2b*. 5'RACE with primer no.17 (Table 2.7) was used to amplify the 5' end of *ROBO2b*. The amplified RT-PCR and RACE sequences are shown in the appendix (Supplementary figure 3 and 4). Full-length cDNA sequences were assembled from cloned cDNA fragments, IMAGE5767984 (gi: 19377431), and Ensembl *ROBO2* sequence. The *ROBO2a* transcript contains 27 exons and is 5920 bp long. *ROBO2b* contains 26 exons and is 6235 bp long.

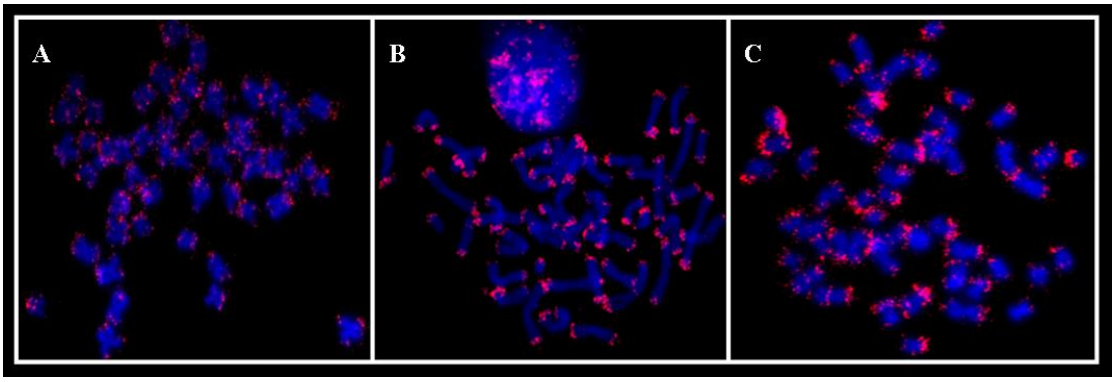



Figure 4.13. Human BAC clones RP11-73I16 (A), RP11-58L10 (B) and RP11-214P4 (C) hybridized to multiple sites in the siamang gibbon (HSY) genome.

### 4.6.2. Expression patterns *of ROBO2* isoforms

In order to study the expression patterns of *ROBO2a* and *ROBO2b*, RT PCR was performed on different human tissue RNAs (using 2 µg each as template for RT) (Figure 4.14C). Primer no.1 (Table 2.7) was used to synthesize the first strand of *ROBO2* cDNA with reverse transcriptase. Subsequent PCRs with primer pairs nos.1 and 2 and nos. 1 and 3, respectively, were used to amplify *ROBO2a* and *ROBO2b* sequences. *ROBO2b* was transcribed in all tissues examined except skeletal muscle. The transcription level in bone marrow, salivary gland, and thyroid gland was much lower than in other tissues. Because the same cDNA templates were used for amplification of *ROBO2a* and *ROBO2b* fragments, the ubiquitous expression of *ROBO2b* served as an internal control. In contrast to *ROBO2b*, *ROBO2a* had a much more restricted transcription pattern (Figure 4.14C). The highest transcription level was observed in fetal brain with significantly lower expression levels in brain, kidney, spleen, testis, and spinal cord. The temporal expression patterns of the two isoforms during human brain development was studied by RT-PCR analysis on human fetal brain RNAs from different development stages (Figure 4.14D). Transcription of both isoforms was detected in all developmental stages without obvious temporal regulation of transcription. Similar to the two isoforms of *ROBO1*, the two *ROBO2* isoforms utilize alternative exons containing different 5'UTRs and translation start sites. The start codon for isoform a is located in exon 2 and that for isoform b in exon 1. Consequently, the two cDNA isoforms encode two protein isoforms with different N-termini. The SignalP program (Bendtsen *et al.*, 2004 and Nielsen *et al.*, 1997) was used to predict the protein sequence. Although different signal peptides are used in the two ROBO2 protein isoforms, the mature proteins differ only in four amio acids at their N-terminal ends. (Figure 4.14 B).

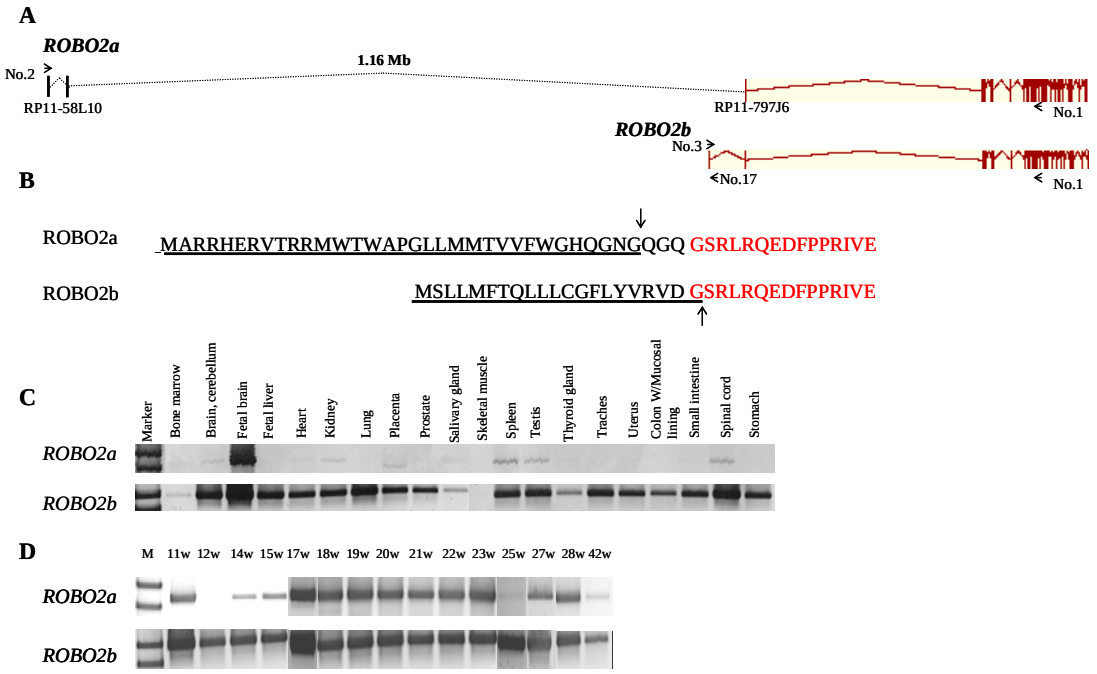



Figure 4.14. Alternative splicing of the *ROBO2* gene generates two differentially expressed transcripts. (A) Genomic organization of the human *ROBO2* gene. The first two exons of isoform a are located in BAC RP11-58L10, which is ~ 1.2 Mb distal to RP11-797J6. RP11-797J6 contains the first two exons of isoform b. The position and direction of primers used for amplification of the two transcripts and the 5' end of isoform b are indicated by arrows. The primer numbers correspond to those in table 2.7. (B) The translated products with the predicted signal peptides are underlined. The vertical arrows indicate the cleavage position. Identical amino acids in both isoforms are marked in red. (C) Transcription pattern of *ROBO2* in human tissues. Isoform b is transcribed in almost all tissues examined. Isoform a is most highly transcribed in fetal brain. (D) Transcription pattern of *ROBO2* in human fetal brain at different developmental stages. Transcription of both isoforms is detected at all stages examined. M: molecular DNA marker. w: week of embryonic/fetal development.

## 4.6.3. Methylation analysis of the promoter region of human *ROBO2a*

Approximately 70-80% of CpG sites in the human genome are methylated. Most of the remaining unmethylated CpG sites are located in CpG islands that are protected from methylation (Antequera and Bird 1993; Bird 1993; Bird 1995). CpG islands are short stretches of DNA, often <1 kb, frequently containing unmethylated CpG dinucleotides. CpG islands tend to mark the 5' ends of genes. Fully methylated CpG islands are found only in the promoters of silenced alleles of selected imprinted autosomal genes (Forne *et al.*, 1997; Reik and Walter 1998) and multiple genes on the inactivated X-chromosomes of females (Heard *et al.*, 1997). Analysis of the sequence surrounding the first exon of *ROBO2a* revealed a CpG island with a GC content >0.5 and a CpG

ratio >0.6 (Figure 4.15). In two human cell lines studied, HELA and HEK293, *ROBO2b* was strongly expressed, whereas *RORO2a* showed an extremely low, if any detectable level of expression (Figure 4.16). In order to find out whether CpG methylation was responsible for transcriptional silencing of *ROBO2a* in these cell lines and, by extrapolation, in many other human tissues, HELA and HEK293 cell cultures were treated with the demethylation agent 5'-aza-2'-deoxycytidine for 2-11 days. The RNA was then isolated from cell pellets and RT PCR was performed to detect *ROBO2a* transcription. The ubiquitously expressed isoform *ROBO2b* was amplified as an internal control. 5'-aza-2'-deoxycytidine treatment did not reactivate *ROBO2a* in the analyzed cell lines (Figure 4.16), indicating that methylation may not be the regulatory mechanism for *ROBO2a* silencing.

tctgctcctg**CG**aact**CG**tttctgg**CG**tggggcttgtg**CG**caccccagac**CG**ggaggagc

**exon 1**

ct**CG**agcctgctcattagtagca**CG**ggcagct**CGC**{**GCG**ctggaggagggaggg**CG**g

aaggacag**CG**ctg**CG**ccaccacc**CG**gaggagggag**CGCG**gtagctgcaggcaggg

gagggagaggaaagaaaaggaagga**CG**gctcccagacagagagtgggagaaac**CG**ggga

gcag**CG**ggagcagcaggtc**CG**gggggagctgttc**CG**ctg**CG**ctgccct**CG**ttattcaca

**CG**ga**CG**ctg**CG**gagcttcccagggctgcttccctgtcccctgggtggaggctgc**CG**tcta

aacctgactccag}

Figure 4.15. The CpG island at exon 1 of *ROBO2a* was identified with the CpG island searcher program "GrailEXP". The CpG island is marked in gray, CpGs are highlighted in bold.

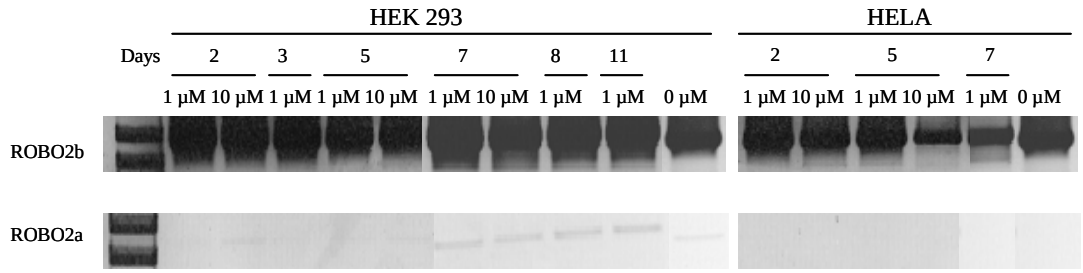



Figure 4.16. *ROBO2* transcripts in AZA-treated HEK293 and HELA cell lines. Cell lines were exposed to two different concentrations (1 µM and 10 µM) of 5-azadeoxycytidine. *ROBO2b* showed strong expression in both cell lines independent of the treatment. Under the same conditions *ROBO2a* showed extremely low, if any expression.

### 4.6.4. Characterization of the mouse homologous of *ROBO2*

Alignment of the two human *ROBO2* isoforms with the mouse genome sequence revealed a similar genomic structure of *Robo2* in the mouse genome (Figure 4.17A). The isoform which occupied a larger genomic region was designated as *MRobo2a* and the smaller one as *MRobo2b*. Several pairs of primers were designed according to the mouse genomic sequence and RT PCRs were performed to amplify cDNA fragments of the mouse *Robo2* isoforms from adult brain RNA (Table 4.7). Full-length cDNA sequences were assembled from overlapping cDNA fragments and Ensembl Gene Robo2 (ENSMUSG00000052516). The expression patterns of the two mouse *Robo2* isoforms were analyzed by RT-PCRs on RNA samples from different adult mouse tissues and whole embryos at different developmental stages (5µg total RNA from hybrid mouse Cast/Ei x B6 as template; Figure 4.17C). Primers nos. 1 and 18 (Table 2.7) amplified a 1531 bp fragment of mouse *MRobo2a*. Primer nos. 1 and 3 amplified a 1465 bp fragment of *MRobo2b*. Similar to human *ROBO2*s, the transcription of *MRobo2a* was only detectable in brain, suggesting that the longer isoform has more specific transcription pattern than *Robo2b*. In contrast to human *ROBO2a* which showed a very low transcription level in adult brain, mouse *Robo2a* was highly transcribed in adult brain. The MRobo2a and b protein sequences predicted by SignalP program (Bendtsen *et al.*, 2004 and Nielsen *et al.*, 1997) revealed that similar to the situation in humans different signal peptides are used by the two isoforms but that the N-terminal of mature proteins differ only in four amino acids.

Table 4.7. Primer pairs and vectors used for cloning of mouse *Robo2*. The positions of primers are indicated in appendix.

| | | Forward primer | | Reverse primer | | Vector |
|---|---|---|---|---|---|---|
| | | No. in table 4.8 | Sequence (5'- 3') | No. in table 4.8 | Sequence (5'- 3') | |
| Fragment 1 | *MRobo2a* | 4 | agc tct gta gct gac agg c | 1 | tgt agc cac aca agt ata agt | pCR®-XL-TOPO® |
| | *MRobo2b* | 3 | gag tct gct gat gtt tac ac | 1 | tgt agc cac aca agt ata agt | pCR®II-TOPO® |
| | Primer walking in fragment 1 | 5 | gat gtg gta gtc gca gct g | 6 | gca acg atc tga tct ctt gg | pCR®II-TOPO® |
| Fragment 2 | *MRobo2* | 7 | act tat act tgt gtg gct aca | 8 | ctc cag tac aac cag atg ct | pCR®II-TOPO® |
| | Primer walking in fragment 2 | 9 | agc aat agc tgg cag aca g | 10 | act aca gag cga atg gct g | pCR®II-TOPO® |
| Fragment 3 | | 11 | agc atc tgg ttg tac tgg ag | 12 | agt ggg cag tgg act gct gt | pCR®II-TOPO® |
| Fragment 4 | *MRobo2* | 13 | aca gca gtc cac tgc cca ct | 14 | gga act cca tgt tat agt ctg | pCR®II-TOPO® |
| | Primer walking in fragment 4 | 15 | tac agc aag ccc agc ttc c | 16 | tac acc agc att atc caa cc | pCR®II-TOPO® |

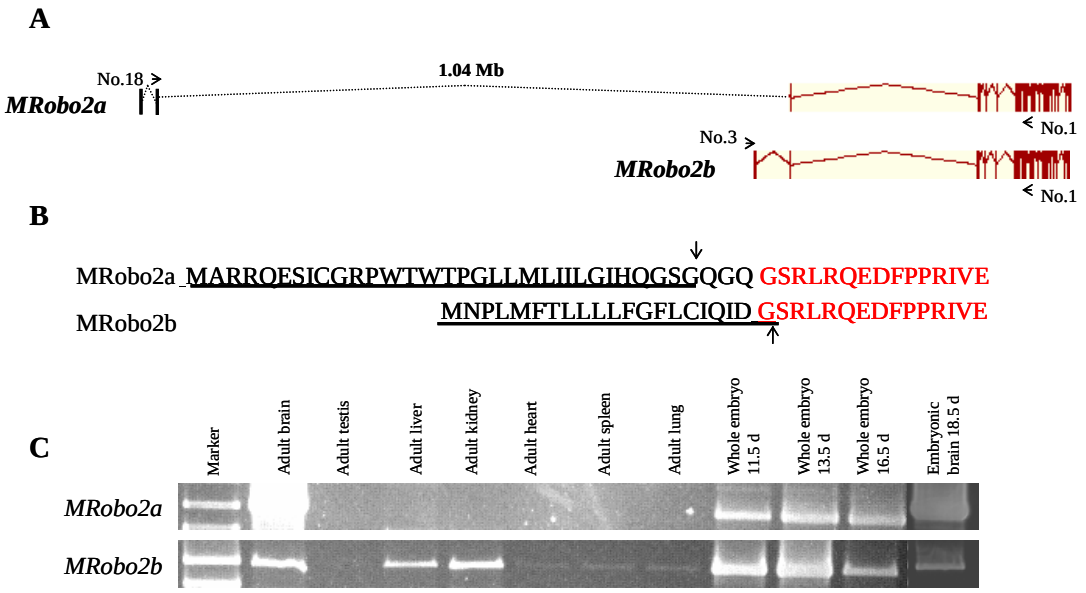



Figure 4.17. (A) Genomic organization of mouse *Robo2*. The first two exons of isoform a are ~1.04 Mb distal to the first exon of isoform b. The position and direction of primers used to amplify the two transcripts are indicated by arrows with the numbers corresponding to table 2.7. (B) The translated proteins with the predicted signal peptides are underlined. Arrows indicate the predicted cleavage position. Identical amino acids in both isoforms are marked in red. (C) Temporal and spatial expression patterns of *ROBOa* and *b* in mouse.

In addition to different 5'UTRs of mouse *MRobo2*, there appear to be multiple alternative splice forms at the 3' end. Alignment of the Ensembl mouse *Robo2* sequence with the cDNA fragment amplified using primers nos.13 and 14 revealed that 273 bp of exon 25 of Ensembl *MRobo2* are not present in the amplified cDNA fragment. Secondly, two different transcripts were amplified with primer pair 11 and 12. Alignment with the Ensembl sequence showed that one transcript contained an additional exon of 126 bp between exons 20 and 21. Altogether, there were at least four variants of each *MRobo2a* and *MRobo2b*, with the longest transcript of 8339 bp containing 29 exons (Figure 4.18 and Table 4.8). The sequences of these eight *MRobo2a* variants are shown in appendix ( from Supplementary figure 5 to figure 12). Consequently, there are also eight mouse ROBO2 protein variants. Multi-alignment analysis of the eight mouse MRobo2 and two human ROBO2 proteins with MultAlign program (data shown in appendix - Supplementary figure 2) showed a high degree of evolutionary conservation at the protein level.

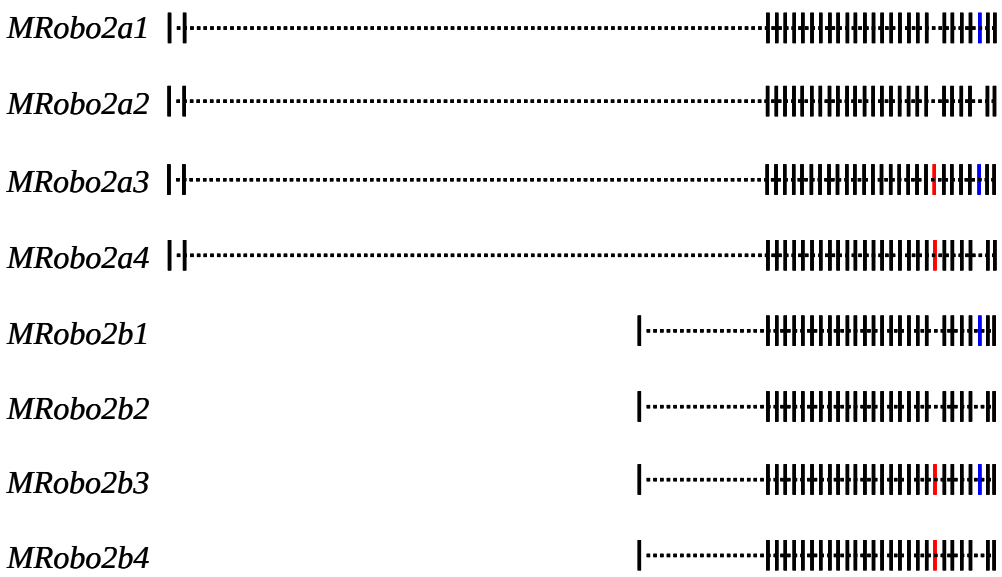



Figure 4.18. Genomic structure of eight mouse *Robo2* variants. The exons are indicatd by vertical bars and the introns by horizontal dotted lines. The additional exon between exons 20 and 21 of Ensembl mouse *Robo2* is marked in red. Exon 25 of Ensembl mouse *Robo2* is marked in blue. Exons and intron lengths are not drawn to scale.

Table 4.8. Variants of mouse *MRobo2*. "+" and "-" indicate presence or absence of a particular exon.

| Isoforms | Variants | Exons | Additional exon (126 bp) | Exon 25 of Ensembl mouse Robo2 (237 bp) | Transcript length (bp) | Start codon - Stop codon (bp)/(exon) | Length of amino acid sequence |
|---|---|---|---|---|---|---|---|
| MRobo2a | 1 | 28 | - | + | 8213 | (236-4693)/(2-28) | 1485 |
| | 2 | 27 | - | - | 7940 | (236-4420)/(2-27) | 1394 |
| | 3 | 29 | + | + | 8339 | (236-4819)/(2-29) | 1527 |
| | 4 | 28 | + | - | 8066 | (236-4546)/(2-28) | 1436 |
| MRobo2b | 1 | 27 | - | + | 7921 | (1-4402)/(1-27) | 1466 |
| | 2 | 26 | - | - | 7648 | (1-4128)/(1-26) | 1375 |
| | 3 | 28 | + | + | 8047 | (1-4527)/(1-28) | 1508 |
| | 4 | 27 | + | - | 7774 | (1-4254)/(1-27) | 1417 |

# 5.1. Genomic structure of evolutionary breakpoint regions

## 5.1.1. Segmental duplications at breakpoint regions

A variety of mutational events including small-scale (single-base pair changes, microsatellite slippage, insertion/deletions) and large-scale events (retrotransposition, chromosomal rearrangements, segmental duplications) are thought to have molded the human genome. Segmental duplications and low-copy repeats (LCRs) play a role in several genomic disorders such as Williams-Beuren syndrome on 7q11.23 and Smith-Magenis syndrome on 17p11.2 (reviewed by Stankiewicz and Lupski, 2002; Shaw & Lupski, 2004). Segmental duplications in the human genome are also present at more than half of the regions of breaks of synteny between the human and mouse genomes (Armengol *et al.,* 2003). Even when excluding pericentromeric and subtelomeric regions which are enriched with segmental duplications, there are still >25% of breaks of synteny associated with duplicated DNA segments of at least 10 kb length (Bailey *et al.,* 2004). Recent high-resolution analyses of evolutionary chromosome rearrangements during higher primate evolution revealed that many evolutionary breakpoints are associated with segmental duplications, for example the telomere-telomere fusion that gave rise to human chromosome 2 (Fan *et al.,* 2002), the translocation involving the gorilla homologs of human chromosome 5 and 17 (Stankiewicz *et al.,* 2001), as well as multiple inversions involving the primate homologs of human chromosome 7, 12, 15 and 18 (Locke *et al.,* 2003; Goidts *et al.,* 2004; Müller *et al.,* 2004; Kehrer-Sawatzki *et al.,* 2005b and Shimada *et al.,* 2005). In this study FISH mapping, PCR mapping and *in silico* analyses revealed segmental duplications in three evolutionary breakpoint regions between human chromosome 3 and Bornean orangutan chromosome 2. Collectively, these results suggest segmental duplications are associated with chromosomal rearrangements during hominoid evolution.

Among hominoid primates, the occurrence and chromosomal distribution of segmental duplications show drastic differences. One example is the 3q21.3 breakpoint cluster region which contains a DNA segment that has been duplicated to at least other 11 loci on 7 different chromosomes in the human genome. In orangutan, this segment highlighted only the subtelomeric regions on p arms of PPY 10 and PPY18, but not the 3q21.3-syntenic site (Figure 4.6 and Table 4.2). Another example is the human BAC clone RP11-58L10, which mapped proximal to the 3p12.3 breakpoint on orangutan chromosome 2 and to multiple other mainly subtelomeric regions (Figure 4.3B and Figure 4.4F). In siamang gibbon, RP11-58L10-homologous sequences were

extensively amplified in the subtelomeric regions of all chromosomes (Figure 4.13). This is in stark contrast to the human genome, where RP11-58L10 mapped only to HSA 3p12.3 and the centromeric regions of two other chromosomes (data shown in appendix - Supplementary figure 1J).

### 5.1.2. Involvement of 7E OR genes in evolutionary chromosome rearrangements

Olfactory receptor (OR) genes comprise the largest gene family in the mammalian genome. The encoded proteins are responsible for odorant binding and discrimination (Buck and Axel, 1991). The 7E subfamily of OR genes has been duplicated intra- and interchromosomally to >30 regions throughout the human genome (Newman and Trask, 2003). Segmental duplications containing 7E OR genes account for approximately 10% of all human OR genes and 50% of the locations where OR genes are found, indicating that they have played an important role in shaping the human OR subgenome (Trask *et al.*, 1998; Glusman *et al.*, 2000a; 2001; Young *et al.*, 2002). Gene conversion among 7E OR genes and sequence exchange between paralogous regions support the idea that highly similar stretches of DNA facilitate interactions between distant loci and, therefore, may predispose to chromosome breakage and rearrangement (Newman and Trask, 2003). Two pairs of 7E OR gene clusters at 4p16 and 8p23, respectively, are thougfht to be responsible for pathological rearrangements in the human genome (Giglio *et al.*, 2001; 2002). Heterozygous 7E OR-related submicroscopic inversions at 4p16 and 8p23, which are are found in 12.5% and 26% of a European population, respectively, may mediate recurrent translocations [i.e. t(4;8)(p16;p23)] and unbalanced rearrangements [i.e. inv dup (8p), der (8)(pter-p23.1::p23.2-pter), and del(8)(p23.1p23.2)] through illegitimate recombination events.

The two evolutionary breakpoint regions at 3p12.3 and 3q21.3 share DNA segments of >110 kb length with >90% sequence similarity, containing four and two OR genes of 7E subfamily respectively (Figure 4.12). FISH mapping on Old World monkey and hominoid chromosomes as well as PCR typing of flow sorted chromosomes demonstrated that an insertion of an approximately 200 kb duplicated 7E OR-containing DNA segment occurred during the 3q21 rearrangement that led to the human and chimpanzee chromosome form, after divergence of the orangutan lineage 12-16 million years ago. In this case, segmental duplication at HSA3q21.3 may have been the result and not the cause of a rearrangement.

Previously, the *myosin light chain kinase* pseudogene, *MYLKP*, was reported to be

closely linked to a cluster of OR pseudogenes in BAC RP11-241K7 and RP11-803B1 at HSA 3p12.3 (Brand-Arpon *et al.,* 1999). By database analysis, the functional *MYLK* gene was located in BAC RP11-9N20. The closest OR gene cluster to *MYLK* was located in BAC CTA-388B5 in the paralogous 3q21.2 segment (Table 4.3) approximately 2 Mb distal of RP11-9N20. During primate evolution a cluster of OR genes was duplicated at multiple genomic sites, including the site near the *MYLK* gene in 3q21. The cluster of *MYLKP* and OR pseudogenes in 3p12.3 is assumed to be the product of an intrachromosomal duplication before the radiation of the great apes. Figure 5.1 presents a model of the evolutionary events forming the paralogous segments on HSA 3. The present day HSA 3q21.3 contig is assumed to represent the ancestral site and most closely resembles a hypothetical ancestral contig I. This ancestral contig was duplicated, inverted and inserted into contig II, which lay approximately 4 Mb apart on the ancestral chromosome. The newly formed contig III corresponds to the present day 3q21.2 contig. The 3q21.1 contig arose through a partial duplication and insertion of contig III sequences at a nearby site in HSA 3q21.1→q21.2. In addition, contig III was duplicated and inverted to the short arm, creating contig IV. This contig evolved into present day HSA 3p12.3. The 7E OR-containing BAC clone RP11-803B1 hybridized to both the 3p12.3- and 3q21.3-syntenic regions in chimpanzee and humans (data shown in appendix - Supplementary figure 1A and B), but it did not react with the 3p12.3-syntenic site in orangutan. This is consistent with the view that the duplicated 7E OR-containing sequences were deleted from the 3p12.3-syntenic site on orangutan chromosome 2 (Figure 4.4 and Table 4.1). However, hybridization to other species is necessary to find out whether the HSA 3p12.3-syntenic segmental duplication was present prior to the divergence of orangutans and African apes and then deleted in the orangutan lineage or whether it was inserted later in evolution during the 3p12.3 rearrangement leading to the present-day chromosome form of humans and African great apes.

7E OR-containing segments are also located at two other evolutionary inversion breakpoints between HSA 7 and homologous hominoid chromosomes (Müller *et al.*, 2004). The 7p22.1 and 7q21.3 breakpoint regions share 24.8 kb with >94% sequence similarity. It is plausible to assume that unequal crossing over between the 7E OR-containing segmental duplications, which are located at the distal sites of the 7p22.1 and 7q21.3 breakpoint contigs, led to the formation of the pericentric inversion in an African ape ancestor.

Using biocomputational methods, Bailey *et al.* (2002a) identified 169 regions in the human genome that are flanked by highly similar segmental duplications and,

therefore, possible hot spots of genomic instability. Interestingly, 7 (>4%) of these supposedly unstable regions contain 7E OR genes (Table 5.1). The 7E OR-containing region at 7p22.1 was involved in evolutionary chromosome rearrangements (Müller *et al.*, 2004), whereas the two pairs of 7E OR gene clusters at 4p16 and 8p23, respectively are responsible for pathological rearrangements (Giglio *et al.*, 2001; 2002).

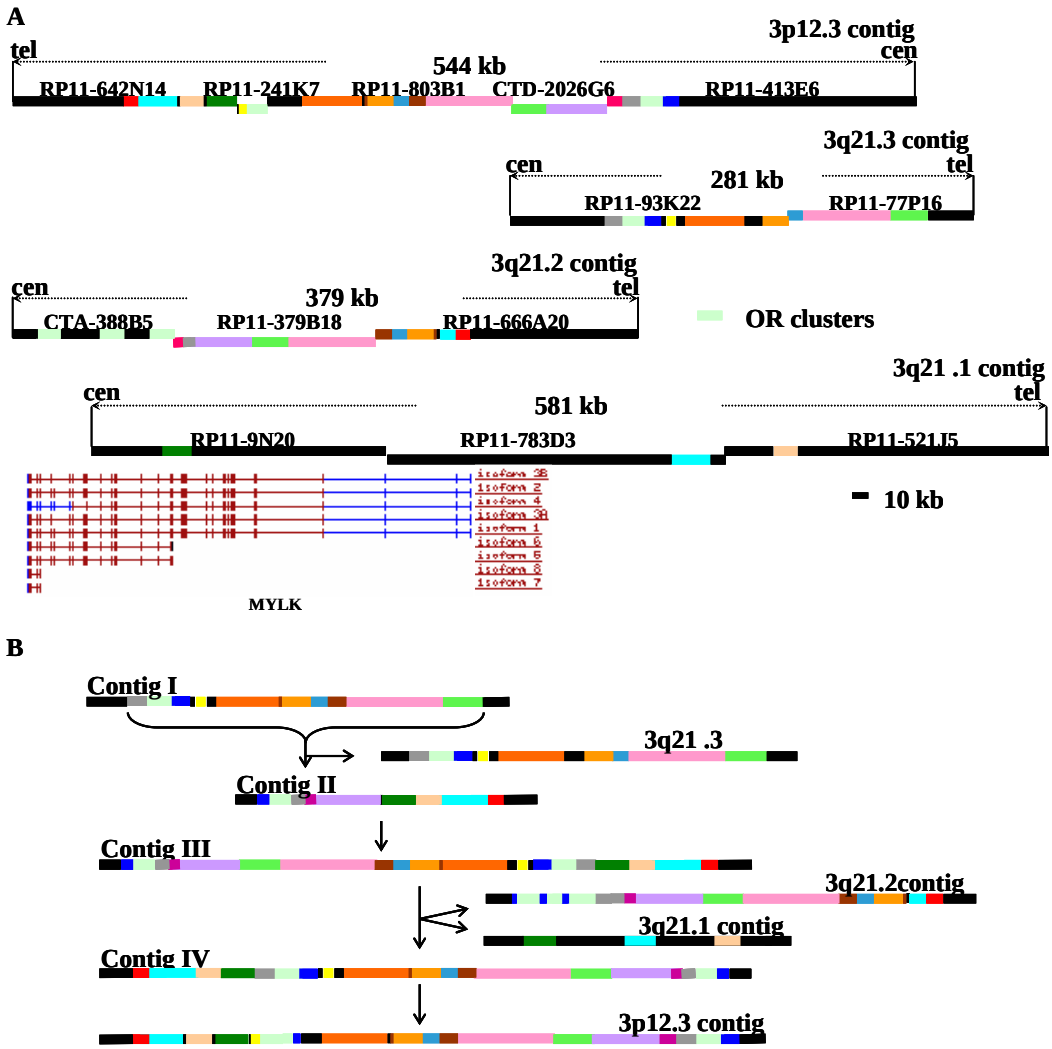



Figure 5.1. (A) Paralogous sequence blocks (>90% sequence similarity) in the HSA 3p12.3, 3q21.3, 3q21.2 and 3q21.1 contigs are indicated by identical colors. Unique sequences are indicated in black. Each uninterrupted bar represents an individual BAC clone. The paralogous segment in the 3p12.3 breakpoint-spanning contig encompasses 5 clones from RP11-642N14 to RP11-413E6. The 152 kb long paralogous segment in 3q21.3 is contained in RP11-93K22 and RP11-77P16. The 169 kb paralogous segment in the 3q21.2 contig is distributed on CTA-388B5, RP11-379B18 and RP11-666A20. The 3q21.1 contig consists of BACs RP11-9N20, RP11-783D3 and RP11-521J5 and contains three smaller paralogous sequence stretches interrupted by unique sequence. The genomic structure of MYLK isoforms is derived from NCBI database. Partial exonic sequence of *MYLK* in RP11-9N20 (dark green) was duplicated to 3p12.3 in RP11-642N14.

(B) Hypothesized evolutionary history of the paralogous sequence blocks on HSA 3. Contigs I-IV represent hypothetical ancestors of the present day regions, contig I being the evolutionarily oldest structure. In this model, the present day 3q21.3 resides at the ancestral site and most closely resembles the ancestral contig I. The 3q21.2 contig is the present day counterpart of a hypothetical contig III, which originated through duplication, inversion and insertion of contig I into contig II approximately 4 Mb proximal to the ancestral site. The 3q21.1 contig represents a partial duplication and insertion of contig III sequences at 5–6 Mb proximal of the ancestral site. The evolutionarily youngest paralogous region at 3p12.3 evolved from a hypothetical contig IV, which is a duplication product of contig III, which was inverted onto the short arm of present day HSA 3 quite far away from the ancestral site.

Table 5.1. Hot spots of genomic instability containing 7E OR genes.

| ***Genomic position*** | ***7E OR genes*** | ***BACs*** |
|---|---|---|
| 2q11.1 (95.6 Mb) | OR7E102P | RP11-440D7 |
| 3q21.2 (126.9-127.1 Mb) | OR7E130P, OR7E29P, OR7E97P, OR7E93P, OR7E53P | CTD-388B5 to RP11-379B18 |
| **4p16.3** (4.2 Mb) **4p16.1** (9.1-9.4 Mb) | OR7E103P, OR7E99P, OR7E43P OR7E84P, OR7E85P, OR7E83P, OR7E86P, OR7E35P | RP11-324I10 RP11-264E23 to RP11-751L19 |
| **7p22.1** (6.6-6.7 Mb) | OR7E39P, OR7E136P, OR7E59P | RP11-577O18 to RP11-740N7 |
| **8p23.1** (7 Mb) **8p23.1** (12.6 Mb) | OR7E125P OR7E15P, OR7E10P, OR7E8P | scb-332A23 RP11-303G3 |
| 9q22.2 (90-90.6Mb) | OR7E31P, OR7E116P, OR7E109P, OR7E108P, | RP11-389K4 to RP11-61N16 |
| 13q21.31 (63.3 Mb) | OR7E104P | RP11-394A14 |

Bailey et al. (2002) identified 169 hot spots of genomic instability flanked by highly similar duplications. Seven of these hot spots (column 1) are endowed with 7E OR genes (column 2). Segmental duplications involved in evolutionary or recurrent pathological chromosome rearrangements are highlighted in bold. The numbers in parentheses indicate the location of the hot spot in the genomic sequence of the respective chromosome. The last column presents the names of the BACs containing or flanking the hot spot.

We have determined the chromosomal synteny of four evolutionary and four pathological 7E OR-containing breakpoint regions in the mouse, rat and chicken genomes (Table 5.2). A break of synteny was identified when a change in orientation or chromosomal location of human-mouse, human-rat and/or human-chicken syntenic segments occurred within 1 Mb of the proximal or distal breakpoint regions. The four breakpoint regions at HSA 3p12.3, 3q21.3, 7p22.1 and 8p23.1 (at 12.6 Mb) are associated with breaks of synteny in all three species analyzed, the breakpoint at HSA 7q21.3 showed a break of synteny in at least one species (chicken). For three breakpoints at HSA 4p16.3, 4p16.1 and 8p23.1 (at 7 Mb), no unique segments with high sequence similarity within 1 Mb of the proximal or distal breakpoint regions could be identified in the current assemblies of the mouse, rat and chicken genomes. Although it is possible that these breakpoints are not associated with breaks of synteny, it is more likely to assume that because of deletions, duplications and/or a high rate of base substitutions these regions are particularly problematic in terms of genome assembly and/or synteny mapping.

Collectively, these data demonstrate the involvement of several 7E OR containing

segmental duplications in human chromosome pathology and evolution.

Table 5.2. 7E OR-containing evolutionary and pathological breakpoint regions and breaks of synteny between the human and the mouse, rat and/or chicken genomes.

| | Location of 7E OR cluster in the human genome | Mouse | Rat | Chicken |
|---|---|---|---|---|
| Evolutionary breakpoints | 3p12.3 (75.5-75.8Mb)[a] RP11-241K7 - RP11-413E6 | 6> - <16[b] | 4> - <11 | 12> - <1 |
| | 3q21.3 (131.2 Mb) RP11-93K22 | 6>- <9 | 4> - <8 | 12> - >2 |
| | 7p22.1 (6.6-6.68 Mb) RP11-577O18 - RP11-740N7 | 5<- >6 | 12> - >4 | 14< - >2 |
| | 7q21.3 (96.5Mb) RP5-1111F22 | 6> - ... (n.d.)[c] | 4> - ... (n.d.) | 2> - >14 |
| Pathological breakpoints | 4p16.3 (4.2 Mb) RP11-324I10 | 5> - ... (n.d.) | 14< - ... (n.d.) | 4< - ... (n.d.) |
| | 4p16.1 (9.1-9.4 Mb) RP11-264E23 - RP11-751L19 | ... - >5 (n.d.) | ... - <14 (n.d.) | ... - <4 (n.d.) |
| | 8p23.1 (7 Mb) scb-332A23 | 8> - ... (n.d.) | 16> - ... (n.d.) | 3> - ... (n.d.) |
| | 8p23.1 (12.6 Mb) RP11-303G3 | 14< - >8 | 15< - <16 | 3> - <4 |

[a] The numbers in parantheses indicate the location of the breakpoint region in the genomic sequence of the respective chromosome. The names of the BACs containing or flanking the breakpoint are written below.

[b] "<" and ">" indicate the direction of syntenic DNA segment(s) in the given mouse, rat or chicken chromosome.

[c] No syntenic mouse, rat and/or chicken segment was identified within 1 Mb of the proximal or distal breakpoint region. Therefore, it could not be determined (n.d.) whether this breakpoint region represents a break of synteny in the respective vertebrate genome.

## 5.1.3. Association of repetitive elements with evolutionary breakpoints

At least 45% of the human genome consists of interspersed repeat sequences including short interspersed elements (SINEs), long interspersed elements (LINEs), long terminal repeat (LTR) retrotransposons, and DNA transposons. SINEs, LINEs and LTRs are propagated by reverse transcription of an RNA intermediate. In contrast, DNA transposons utilize a direct “cut-and-paste” mechanism of DNA sequence. The dynamic human genome has been bombarded with a variety of repeat elements

through successive waves of retrotransposition during the last 60 million years of evolution (Smit 1999; Deininger and Batzer, 2002). Large-scale comparison of primate genomic sequences also revealed a dramatic difference in the retrotranspositon rate, leading to a 15%-20% expansion of human genome size during the last 50 million years, 90% of it due to new retrotransposon insertions (Liu *et al.,* 2003). *Alu* elements, a member of SINEs and the most abundant class of interspersed repeat elements, play a role in mediating some recurrent chromosomal aberrations in tumors (reviewed by Kolomietz *et al.,* 2002) and the pericentric inversion of the chimpanzee homolog of HSA 17 (Kehrer-Sawatzki *et al.,* 2002). The breakpoint regions between human chromosome 19 and the mouse genome are also associated with high concentrations of tandemly organized L1 repeats and LTRs (Dehal *et al.,* 2001). Similarly, the enrichment of the breakpoint contigs at human chromosome 3 with LTRs and other repetitive elements (Table 4.5) may reflect an increased transposition capacity, which seems to be a hallmark of unstable regions.

Because a high percentage (27%) of segmental duplications within the human genome terminate within an *Alu* short interspersed repeat sequence, it has been proposed that Alu-Alu-mediated recombination events account for the expansion of segmental duplications in the hominoid genome (Bailey *et al.,* 2003). Consistent with this *Alu* transposition model for segmental duplications, the 3q21.3 breakpoint cluster region (from 93K22v to 77P16i) is composed of 16.8% *Alu* elements (Figure 4.11 C & D).

## 5.2. Hot spots for evolutionary chromosome rearrangements

There are two models of chromosomal evolution: the random chromosome breakage model and the fragile site breakage model. Ohno (1973) postulated over 30 years ago that evolutionary breakpoints occur randomly throughout the genome. Nadeau and Taylor (1984) found the number and distribution of conserved segments (i.e., syntenic segments with preserved gene order) between human and mouse to be consistent with a random breakage model. At a gross level of resolution, subsequent comparative mapping and sequencing studies among vertebrate species have upheld the random nature of evolutionary rearrangements (Aparicio *et al.,* 2002; The International Human Genome Sequencing Consortium, 2001; Mouse genome sequencing 2002; reviewed by Eichler and Sankoff, 2003). Complete sequencing of a growing number of genomes has confirmed the extensive synteny conservation among vertebrates. However, the high density of markers in the complete genomic sequence revealed many more intrachromosomal rearrangements than previously predicted and resulted in a more

complicated view of chromosomal evolution. Small local inversions appear to predominate in many eukaryotic lineages (Huynen *et al.*, 2001; Seoighe *et al.*, 2000; Kellis *et al.*, 2003). Comparison of the finished genome of *Saccharomyces cerevisiae* and the shotgun sequence from *Candida albicans* provided evidence for approximately 1100 single-gene inversions since divergence of these two species 140 to 330 million years ago. The abundance of small inversions implies many pairs of closely spaced breakpoints and does not fit the random breakage model. The high-resolution comparative analysis of the human and mouse genomes also showed a large number of very short “hidden” synteny blocks that were not detectable by comparative gene mapping approaches and an extraordinary density of breakpoints within particular regions of the genome. Pevzner and Tesler (2003) suggested an alternative model for chromosomal evolution, tentatively called the “fragile site breakage model”. An abundance of primate-specific segmental duplications was observed at the breaks of synteny between the human and mouse genome and this association is highly significant, when compared to a simulated random-breakage model. This supports the fragile site model of chromosomal evolution according to which specific regions within the mammalian genome predispose to both small-scale duplications and large-scale evolutionary rearrangements (Armengol *et al.*, 2003; Bailey *et al.*, 2004).
FISH analyses with human YAC clones have shown recurrent evolutionary breakpoints within relatively small (one megabase) intervals in several cases during primate evolution (Table 5.3). Recent nucleotide sequence comparison of the rearrangements between human chromosome 12 q arm and the homologous chimpanzee and gorilla chromosomes indicates that the breakpoints in the two great apes are not identical but only separated by 1.9 kb (Shimada *et al.*, 2005). In this study, high-resolution FISH and PCR mapping results in great apes, siamang gibbon and Old World monkey revealed independent evolutionary breakpoints in an 230-kb DNA segment syntenic to HSA 3q21.3 during hominoid evolution (Figure 4.9 and 4.11). Additional evidence for the nonrandom breakage model of genome evolution comes from comparative analyses of the human, mouse, rat and chicken genomic sequences. All three evolutionary breakpoint contigs between HSA 3 and PPY 2 are in regions of breaks of synteny between the human, rodent, and/or avian genomes (Figure 4.10) indicating that these contigs are hot spots for evolutionary chromosome rearrangements.

Table 5.3. Reuse of evolutionary breakpoint regions: independent breakpoints within one megabase intervals.

| ***Chromosome rearrangement*** | | ***Breakpoint spanning YAC*** |
|---|---|---|
| Pericentric inversion | HSA 12p and PTR XII | 934h5 ( Nickerson and Nelson, 1998) |
| | HSA 12p and GGO XII | |
| Pericentric inversion | HSA 9 and PTR IX | 945f5 (Montefalcone *et al.,* 1999) |
| | HSA 9 and CJA IX | |
| Pericentric inversion | HSA 4q and PTR IVq | 695h10 (Marzella *et al.,* 2000) |
| | HSA 4q and MML IVq | |

HSA: *Homo sapiens,* CJA: *Callithrix jacchus,* GGO: *Gorilla gorilla,* MML: *Macaca mulatta,* PTR: *Pan troglodytes*.

Somatic rearrangements of the short arm of chromosome 3 have been observed in a variety of tumors. The homozygous deletion in a breast cancer cell line HCC38 lies between the markers D3S2537 and D3S2527 in 3p12.3 amd 3p12.2, respectively (Sundaresan *et al.,* 1998). A high density of tumor breakpoints occurs in the 3p12 region from marker D3S1284 to D3S1577 (from 73.0 Mb to 80.4 Mb) representing either deletions and/or hot spots of mitotic recombination (Wistuba *et al.,* 2000). This region covers ~7.4 Mb of the ~10.7 Mb U2020 homozygous deletion region in small cell lung cancers (SCLC) (Rabbitts *et al.,* 1990). This highly unstable region shows frequent allele loss in lung cancers after smoking exposure. The interval from D3S1284 to D3S1511 also shows frequent allele loss in breast carcinomas and precancerous lesions (Maitra *et al.,* 2001). Intriguingly, this region also contains the evolutionary breakpoint contig at HSA 3p12.3 analyzed in this study (Figure 5.2). The coincidence of evolutionarily unstable regions with tumor-associated deletions was also reported for other regions on human chromosome 3p (Kost-Alimova *et al.,* 2003). Two TATAGA repeats at 74.1 Mb and 77.7 Mb respectively, in the 3p12 tumor-associated unstable region (Kost-Alimova *et al.,* 2003; Figure 5.2) flank the evolutionary breakpoint contig. It is conceivable that the length of these repeats is polymorphic in the human population. Repeat expansion might participate in rearrangements during evolution and tumor development. Collectively, these data suggest that human chromosome 3p12.3 is a hot spot for both evolutionary and pathological chromosomal rearrangements.

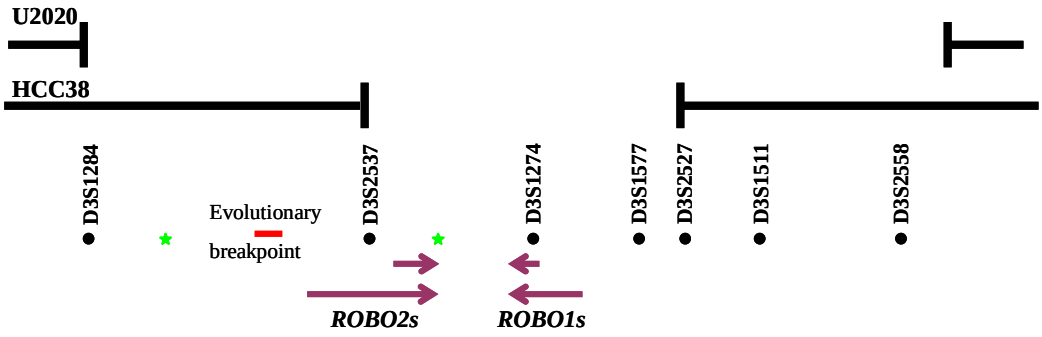



Figure 5.2. Genomic unstable region at HSA 3p12.3. The U2020 deletion region in small lung cancers and the deletion in breast cancer cell line HCC38 are indicated by the gaps between horizontal black lines. The evolutionary breakpoint region is indicated by the red line. STS markers are indicated by black dots and the respective names. Two green stars indicate the TATAGA repeats. Purple arrows indicate the genomic position and transcription direction of *ROBO1* and *ROBO2* isoforms.

## 5.3. Genes at evolutionary breakpoint regions

Cytogenetic and molecular analyses of evolutionary breakpoint regions can provide new insights into the possible effects of evolutionary rearrangements on genome function. It is possible that evolutionary chromosome rearrangements interfere with genome function either directly, by disrupting a gene (s) or indirectly by (in) activating a closely juxtaposed gene(s). Several recent studies on chromosome rearrangements during higher primate evolution argue in favor of the notion that evolutionary breakpoint regions are located within gene poor areas which are rich in repetitive elements and/or low copy repeats (segmental duplications). “Fusion genes” or “truncated genes” which are frequently observed in chromosome rearrangements in cancer seem to be very rare or non-existent in chromosome evolution. So far, four genes, *USP14* and *HSRTSBETA* at the inversion breakpoint between HSA18p11.3 and the homologous chimpanzee chromosome (Goidts *et al.*, 2004; Caceres *et al.*, 2003 & Karaman *et al.*, 2003), as well as *C4orf12* and *WDFY3* at the inversion breakpoint between HSA 4q21.3 and the chimpanzee genome (Kehrer-Sawatzki *et al.*, 2005a) were reported to have different expression level between human and chimpanzee. However, an up-regulation of USP14 was also found in human and gorilla fibroblasts, compared to bonobo (Karaman *et al.*, 2003). Because the pericentric inversion does not exist in gorilla, it is plausible that factors other than the chromosome rearrangement account for the USP14 expression differences among primates. Searching orangutan EST database (BLAST in http://mipc2.gsf.de/proj/cDNA/pest) identified seven orangutan EST and cDNA clones which can be aligned with the HSA 3p25.1 and 3p12.3 breakpoint regions (Table 5.4). Comparative expression analyses in

orangutan and primates without 3p25.1 and/or 3p12.3 rearrangement(s) as well as sequence comparisons of gene regulatory regions are required to elucidate the relationship between evolutionary rearrangements and changed gene expression patterns.

The sequence similarity among paralogous regions in the HSA 3p12.3 and 3q21.3 breakpoint contigs in the human genome makes it difficult and sometimes impossible to map EST and mRNA sequences unequivocally to a specific genomic position. Six genes at HSA 3p12.3, LOC339879, LOC440965, LOC285296, LOC391552, LOC285299 and LOC401074 display ambiguous or conflicting alignments of mRNAs or mRNAs plus ESTs to a genome position.

The analyzed breakpoints do not disrupt the genomic structure of any bona fide genes. Experiments towards confirmation of the gene model LOC440978 will be particularly interesting because the genomic region of this gene spans the distal boundary of the HSA 3q21.3-syntenic evolutionary breakpoint. This gene is most likely the result of an insertion of a ~230 kb DNA segment into the human/African ape chromosome form. This gene is predicted to encode a protein similar to alpha 3 type VI collagen isoform 2 precursor. If this gene can be validated, subsequent functional analysis could provide the first example of a hominoid chromosome rearrangement directly influencing gene function.

Table 5.4. Orangutan EST clones aligned to evolutionary breakpoint regions on human chromosome 3.

| ***EST clones or cDNAs*** | | ***Locus on human chrom. 3 (bp)*** | ***Corresponding UniGene*** | ***Gene*** |
|---|---|---|---|---|
| 3p25.1 (from distal to proximal) | DKFZppcor1_3e19_r1 | 14963616-14963981 | Hs.517821 | IMAGE:5278517 |
| | DKFZppcor1_32c8_r1 | 15062788-15063275 | Hs.128201 | MRPS25 |
| | DKFZppcor1_11e11_r1 | 15063925-15064487 | Hs.128201 | MRPS25 |
| | DKFZppcor1_44a23_r1 | 15086583-15115659 | Hs.475565 | ZFYVE20 |
| 3p12.3 (from distal to proximal) | DKFZppkid1_38h16_r1 | 75634381-75635737 | Hs.274541 | LOC285299 |
| | DKFZppcor1_26l16_r1 | 75708534-75709114 | Hs.543085 | LOC377064 |
| | DKFZP469C036 | 75709147-75755567 | Hs.543085 | LOC377064 |

## 5.4. Roundabout gene family and *ROBO2* isoforms

The *roundabout* (*robo*) gene was first identified in *Drosophila* in a large-scale screen for genes that control the decision of axons to cross the central nervous system midline. In *robo* mutants, too many axons cross and recross the midline (Seeger *et al.,* 1993). The gene encodes an axon guidance receptor that defines a novel subfamily of immunoglobulin superfamily proteins that are highly conserved from fruit fly to man (Kidd *et al.,* 1998). Biochemical purification and genetic linkage analysis identified *slit*, a large extracellular matrix protein, as ligand for *robo* (Kidd *et al.,* 1999). The *slit* and *robo* genes determine the neuronal patterning of commissural neurons crossing the midline in *Drosophila*.

To date, four mammalian *Robo* homologs have been identified: *Robo1/DUTT1*, *Robo2*, *Robo3*, and *magic roundabout* (*Robo4*). *Robo4* is the only *Robo* family member expressed in primary endothelial cells and was identified based on its differential expression in mutant mice with defects in vascular sprouting (Park *et al.,* 2003). It is specifically expressed in the vascular endothelium during murine embryonic development and may provide a chemorepulsive cue to migrating endothelial cells during angiogenesis.

All *Robo* genes except *Robo4*, have five extracellular immunglobulin (Ig) domains and three fibronectin domains, and function as transmembrane proteins at cell surfaces (Figure.5.3B). There is an evolutionarily conserved chromosomal colocalization of *Robo* family members from zebrafish to human (Figure.5.3A). The organization of Robo 1 and 2 and of Robo 3 and 4 gene pairs suggests that the *Robo* genes evolved through a series of gene duplication events.

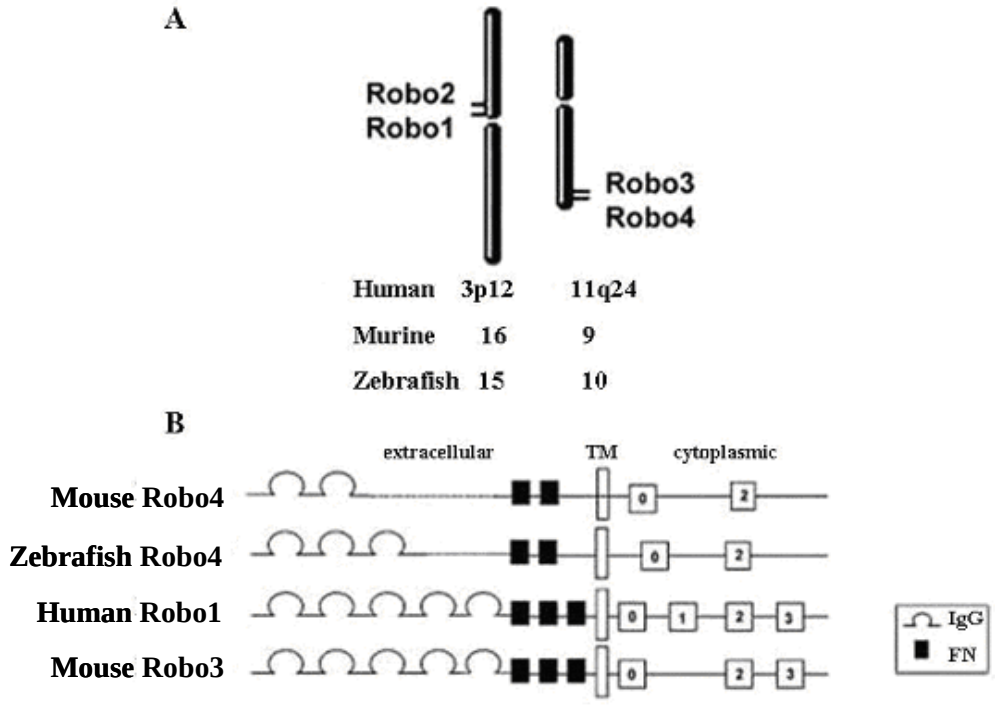



Figure 5.3. (A) *ROBO1* and *ROBO2* are closely linked on human chromosome 3, whereas *ROBO3* and *ROBO4* are juxtaposed on chromosome 11. This paired organization has been conserved in mice and zebrafish. (B) Comparison of the domain organization of some Robo family members. In all other species, members of the *Roundabout* family have five immunoglobulin (IgG) and three fibronectin (FN) domains in the extracellular region. In contrast, murine *Robo4* encodes two IgG and two FN domains. Human *ROBO4* also encodes two IgG and two FN domains (data not shown), zebrafish *Robo4* has three IgG and two FN domains. There is also evolutionary conservation of specific domains within the cytoplasmic regions of *Robo4* (CC0 and CC2) and other *Robo* family members. (Adapted from Park *et al.*, 2003)

In mouse, a low level of Robo1 and Robo2 protein expression was detected in the initial crossing of the midline, while a high level of expression was found in the postcrossing fibers preventing recrossing (Sundaresan *et al.*, 2004). *Robo1* and *Robo2* single mutants showed guidance defects that reveal a role for these two receptors in guiding commissural axons to different positions within the ventral and lateral funiculi (Long *et al.*, 2004). It has been proposed that *Robo1* regulates midline crossing while *Robo2* controls the extension of commissural axons away from the floor plate in the contralateral neural tube. The interaction of *Robo* and *Slit* with other midline signaling pair [Netrin and it's receptor DCC (Deleted in Colon Cancer)] and their role in guiding axons are illustrated in figure 5.4.

Interestingly, *Slit* and *Robo* may also play a role in mouse lung and kidney development. Mice with a targeted deletion of *Dutt1/Robo1* display lung pathology (Clark *et al.*, 2002). *Robo2* is expressed in mesenchymal cells immediately adjacent to large airways, whereas *Slit3* expression predominates in mesenchyme remote from

epithelium. These genes may direct the functional organization and differentiation of fetal lung mesenchyme (Greenberg *et al.,* 2004). Mouse mutants lacking either Slit2 or its receptor Robo2 develop supernumerary ureteric buds that remain inappropriately connected to the nephric duct. In addition, expression of *Gdnf*, which is secreted by the metanephric mesenchyme and elicits ureteric bud formation, is inappropriately maintained in anterior nephrogenic mesenchyme in these mutants. It is assumed that Slit2/Robo2 signaling restricts the extent of the Gdnf expression domain, thereby precisely positioning the site of kidney induction (Grieshammer *et al.*, 2004).

*Robo3* was first identified in a screen for upregulated genes in *retinoblastoma susceptibility* (*RB*)-gene deficient mice. It is expressed in mouse hindbrain and spinal cord. The expression pattern of *Robo3* (*Rig-1*) in the spinal cord overlaps with, but is distinct from the *Robo1* and *Robo2* expression patterns (Camurri *et al.,* 2004). *Robo3*-deficient mice die soon after birth (because of a failure to wean) and show axons that fail to cross the midline (Sabatier *et al.,* 2003). The human *ROBO3* is expressed in fetal human brain and contains multiple alternative splice forms. Mutations in the human *ROBO3* genes were identified in patients affected with horizontal gaze palsy with progressive scoliosis and hindbrain dysplasia (HGPPS) (Jen *et al.*, 2004). The two major nerve tracts, the motor corticospinal tract and the dorsal somatosensory tract, fail to cross the midline in the hindbrain of these patients. Like its murine homolog Robo3, but unlike other Robo proteins, human ROBO3 is required for hindbrain axon midline crossing.

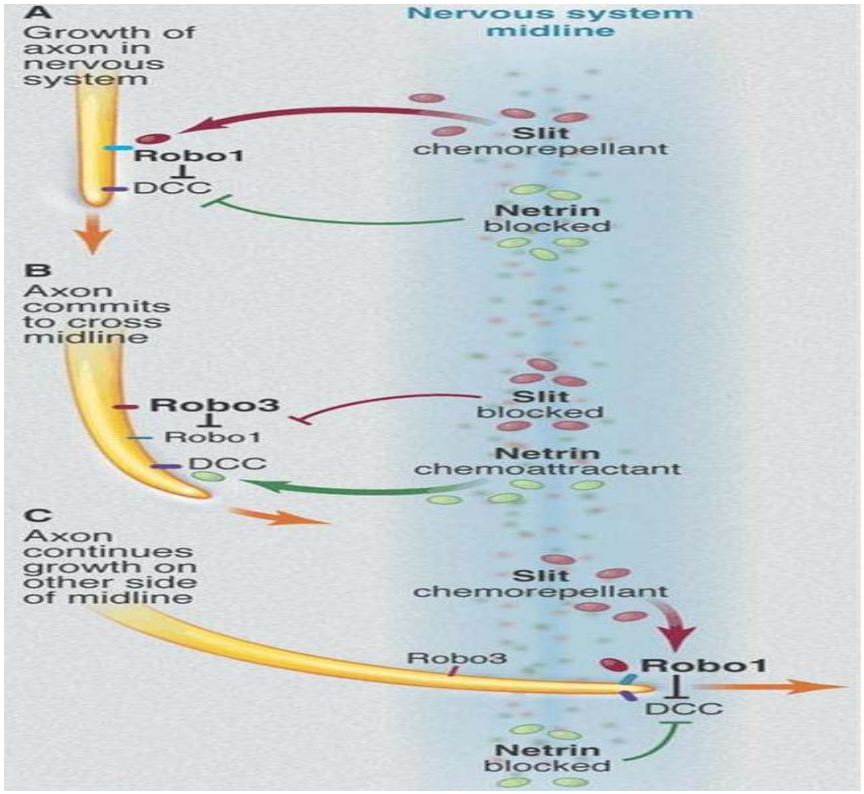



Figure 5.4. (A) As the growing axon tip extends towards its target site in the developing brain, Slit protein produced by the midline repulses the growing axon by binding to the Robo1 receptor expressed in the axon tip. In addition, the chemoattractant effects of Netrin are blocked by interaction of Robo1 with the Netrin receptor DCC (Deleted in Colon Cancer). (B) As the axon tip continues to grow and commits to crossing the midline, Robo3 is strongly expressed. Robo3 releases Robo1's block of the DCC receptor, enabling Netrin to bind to DCC and it acts as a chemoattractant for the growing axon tip. (C) When the axon tip has crossed the midline, Robo1 expression increases markedly and Robo3 is down-regulated. This allows Slit to act as a chemorepellent for the growing axon tip, forcing it out of the midline. The chemoattractant effect of Netrin is blocked again by the interaction of Robo1 with DCC, preventing the axon tip from recrossing the midline. (From Woods 2004)

Human *ROBO1* was first cloned as a potential tumor suppressor gene from the lung cancer tumour suppressor gene region 2 (LCTSGR2 or U2020) at chromosome 3p12 (Figure 5.2) and was first named *deleted in U twenty twenty* (*DUTT1*) (Sundaresan *et al.*, 1998). This gene was independently isolated as human homolog of the *Drosophila robo* gene and entered into GenBank as *H-robo1* (Kidd *et al.*, 1998). Alignment of the two gene sequences and analysis of the draft human genome sequence revealed that *DUTT1* results from alternate splicing of *H-robo1*. The two cDNAs differ only in sequences corresponding to their signal peptides which are cleaved in the mature protein and in their 5' untranslated regions. In this light, *H-robo1* gene is now considered as *ROBO1* isoform a and *DUTT1* as *ROBO1* isoform b. The *ROBO1* isoform a transcript is longer and, following cleavage of the signal peptide, contains 32

unique amino acids at the N-terminus of the mature protein. RT-PCR experiments showed that the mouse *Robo1* isoform b is more widely expressed during development and is frequently the only detectable *Robo1* transcript in adult tissues. Transcription of mouse *Robo1* isoform a was detected in all examined embryonic tissues but only in adult brain, kidney and eye (Clark *et al.,* 2002).

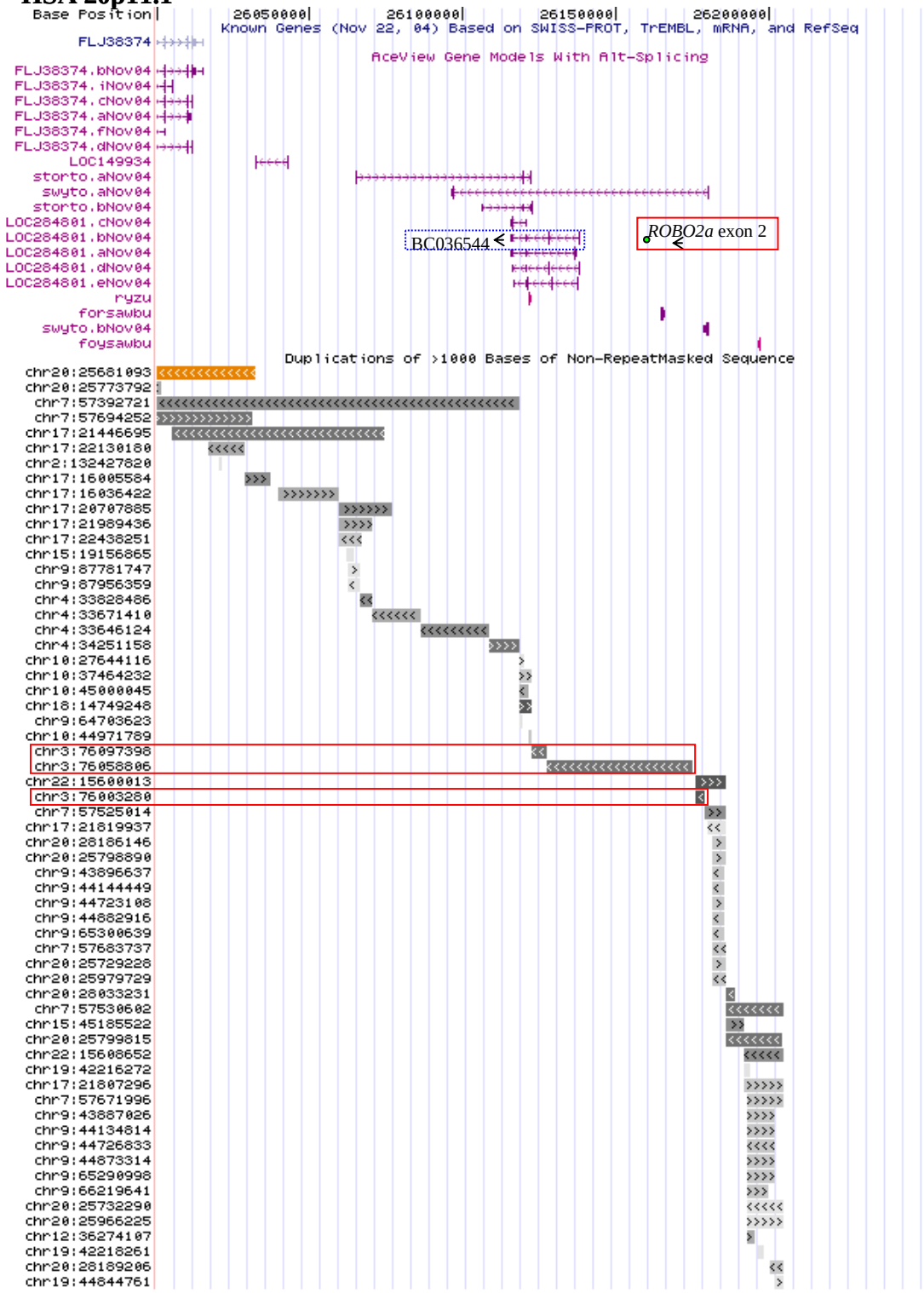
In this study, a new isoform of the human *ROBO2* gene was identified. Similar to the two isoforms of human *ROBO1,* the cDNAs of the two *ROBO2* isoforms differ at the 5' region. By RT-PCR experiments, *ROBO2* isoform b showed a more or less ubiquitous expression, whereas *ROBO2* isoform a was highly expressed only in human fetal brain with lower expression in spinal cord, spleen and testis (Figure 4.14). Considering its function in axon guidance and its abundance in human fetal brain, it is plausible to assume that *ROBO2* isoform a plays an important role in human fetal brain development and is a candidate gene for mental retardation. In 7 of 18 patients with paracentric inversions of human chromosomes (Fryns *et al.,* 1986), an inv(3)(p13p25) has been described with the 3p13 breakpoint in close proximity to the evolutionary breakpoint at HSA 3p12.3. Intriguingly four of these 7 inv(3) patients were reported to suffer form mental retardation and congenital malformations. High-resolution FISH analyses could provide important information on the molecular causes of the mental retardation in these inv(3) patients and the role of *ROBO2* in development of the central nervous system. Because *ROBO1 and ROBO2* are located in chromosomal regions that are unstable in many tumor types (Figure 5.2), both genes are assumed to function as tumor suppressors in addition to their known function in axon guidance. Tumor specific promoter methylation of *ROBO1b* was found in human cancers (Dallol *et al.,* 2002). Further functional analysis of ROBO1 and ROBO2 will help to better understand their role in tumorigenesis.

Two *Robo2* isoforms were also identified in mouse (Figure. 4.17). Their conserved genomic structure and expression patterns in mouse and humans argue in favor of the notion that the two isoforms have different functions and are under selective pressure. The different signal peptides of *ROBO2* isoforms a and b may alter the mRNA half-lives and/or cellular location. All *Robo2* knock-out mice that have been analyzed so far are deficient for both isoforms. It will be interesting to construct isoform-specific knockout mouse strains with mutations in the different 5' regions of the two isoforms and to carefully compare their phenotypes. These experiments will reveal whether or not the two isoforms are functionally redundant and can compensate each other.

In this study, human ROBO2 isoform a was identified by FISH mapping and *in silico*

analyses of the evolutionary chromosome 3p12.3 rearrangement, involving intragenomic duplication of BAC RP11-58L10 sequences. In the human genome, sequences paralogous to RP11-58L10 reside in the centromeric transition regions at HSA 20p11.1 and 22q11.1. The duplicons in these centromeric regions are flanked by duplicated segments that originated from other human chromosomes. The duplicated segments at HSA 22q11.1 and HSA 20p11.1 contain exon 1 and exon 2 of *ROBO2a*, respectively (Figure 5.5). Recent detailed analyses of the structure and assembly of all human pericentromeric regions showed that at least 30% of the centromeric transition regions originated from euchromatic gene-containing DNA segments that were duplicatively transposed towards pericentromeric regions at a rate of six to seven events per million years during primate evolution. This process has led to the formation of a minimum of 28 new transcripts by exon exaptation and exon shuffling (She *et al.*, 2004). Among these new transcripts mRNA BC036544 at 20p11.1 and AK001299 at 22q11.1 are particular interesting because they are in close proximity of the duplicated *ROBO2a* exons and BC036544 has the same transcription direction. Besides these two transcripts, *von Willebrand factor pseudogene* (*VWFP*) at 22q1.1 is located proximal to a duplicated *ROBO2a* exon 1 and has the same transcription direction. It is possible that the duplicated exons of *ROBO2a* are fused to these transcripts and, thus, generate, new transcripts, which may provide new functions for human brain development. RP11-58L10 also hybridized to multiple sites in orangutan and gibbon genomes. Therefore, it is even more likely that *Robo2* exons contributed to the formation of new chimeric transcripts in the orangutan and gibbon lineages. The widespread amplification/duplication of *Robo2a* exons during hominoid evolution may facilitate the rapid evolution of new transcripts which are highly expressed in the developing (fetal) brain.

**HSA 20p11.1**




(continued on next page)

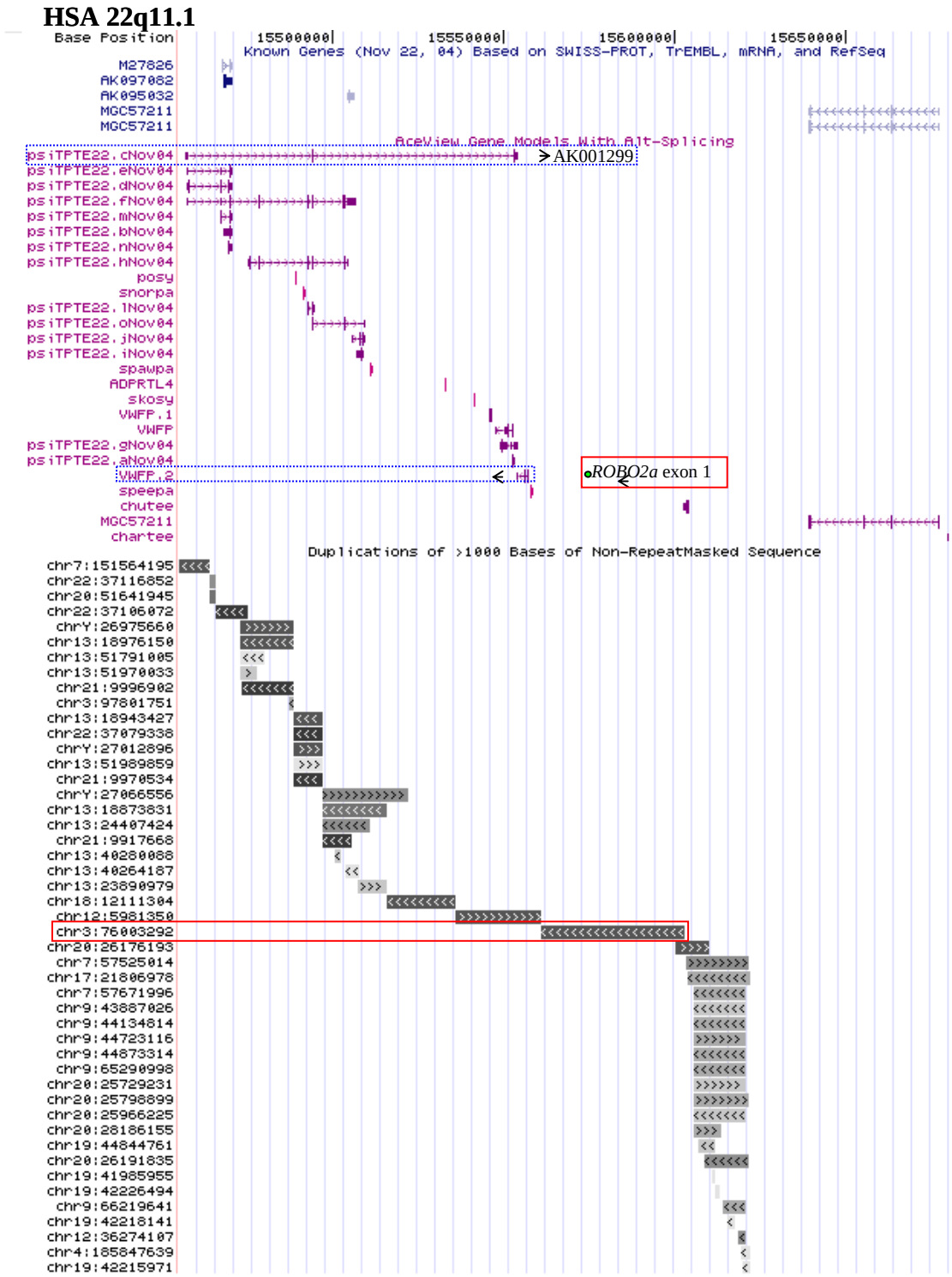
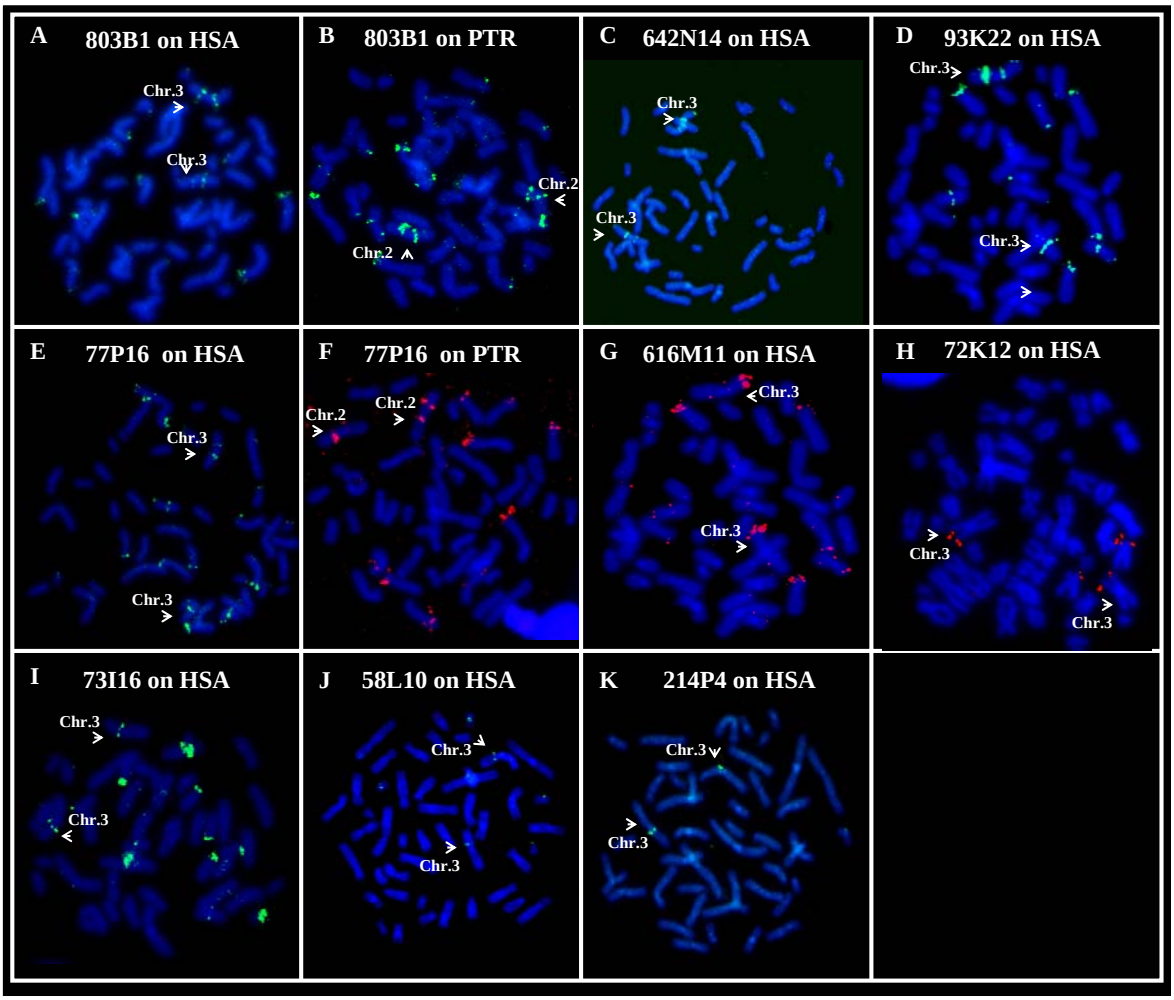


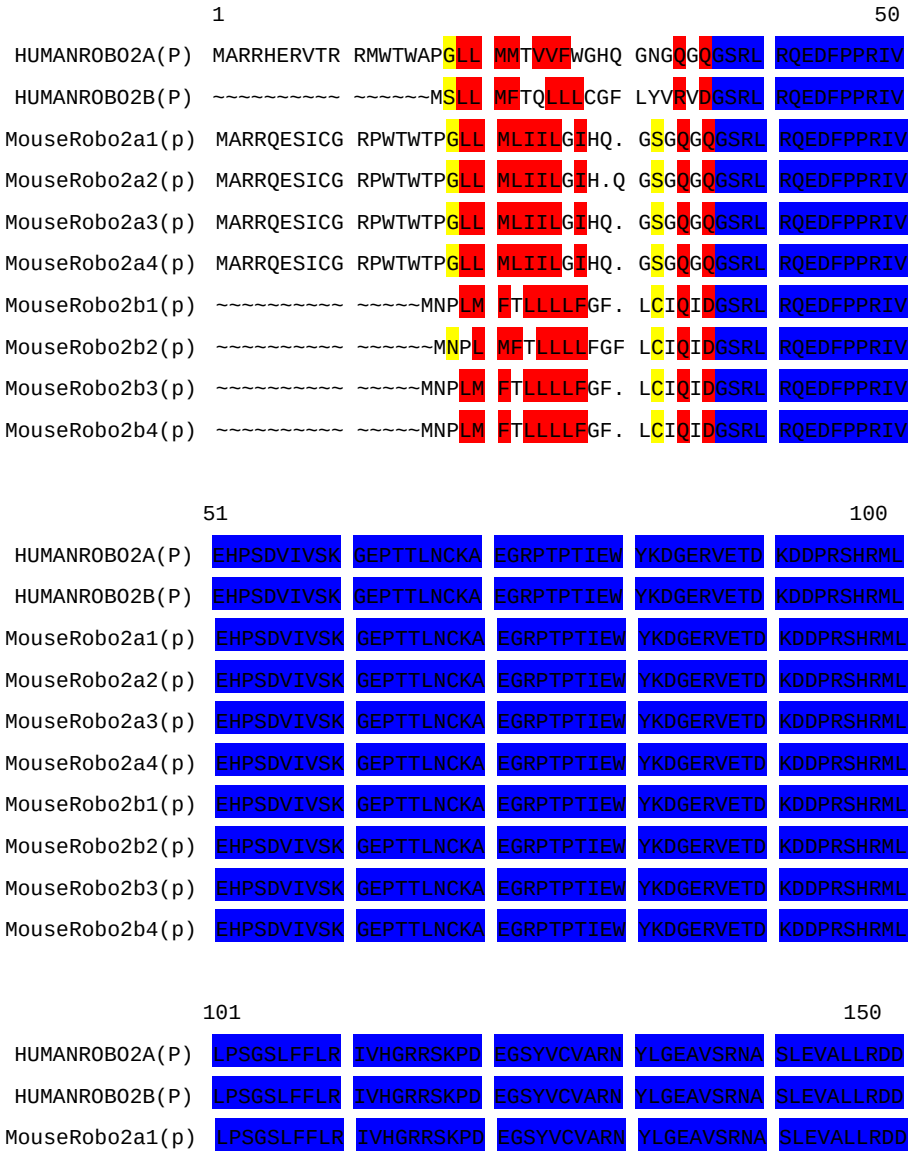

Figure 5.5. Duplication of RP11-58L10 sequences at the pericentromeric regions of chromosome 20p11.1 and 22q11.1 (human genome build 35, May 2004, Chr.3: 76007019-76160982). The duplicated segments are framed in red. The duplicated *ROBO2a* exons are indicated with green cycles. Two new transcripts identified by She *et al.* (2004), BC036544 and AK001299, are framed in blue. The *VWFP* pseudogene which is proximal to duplicated *ROBO2a* exon 1 is also framed in blue. Arrows indicate the direction of transcription. Gene contents and segmental duplications are obtained from UCSC genome browser.

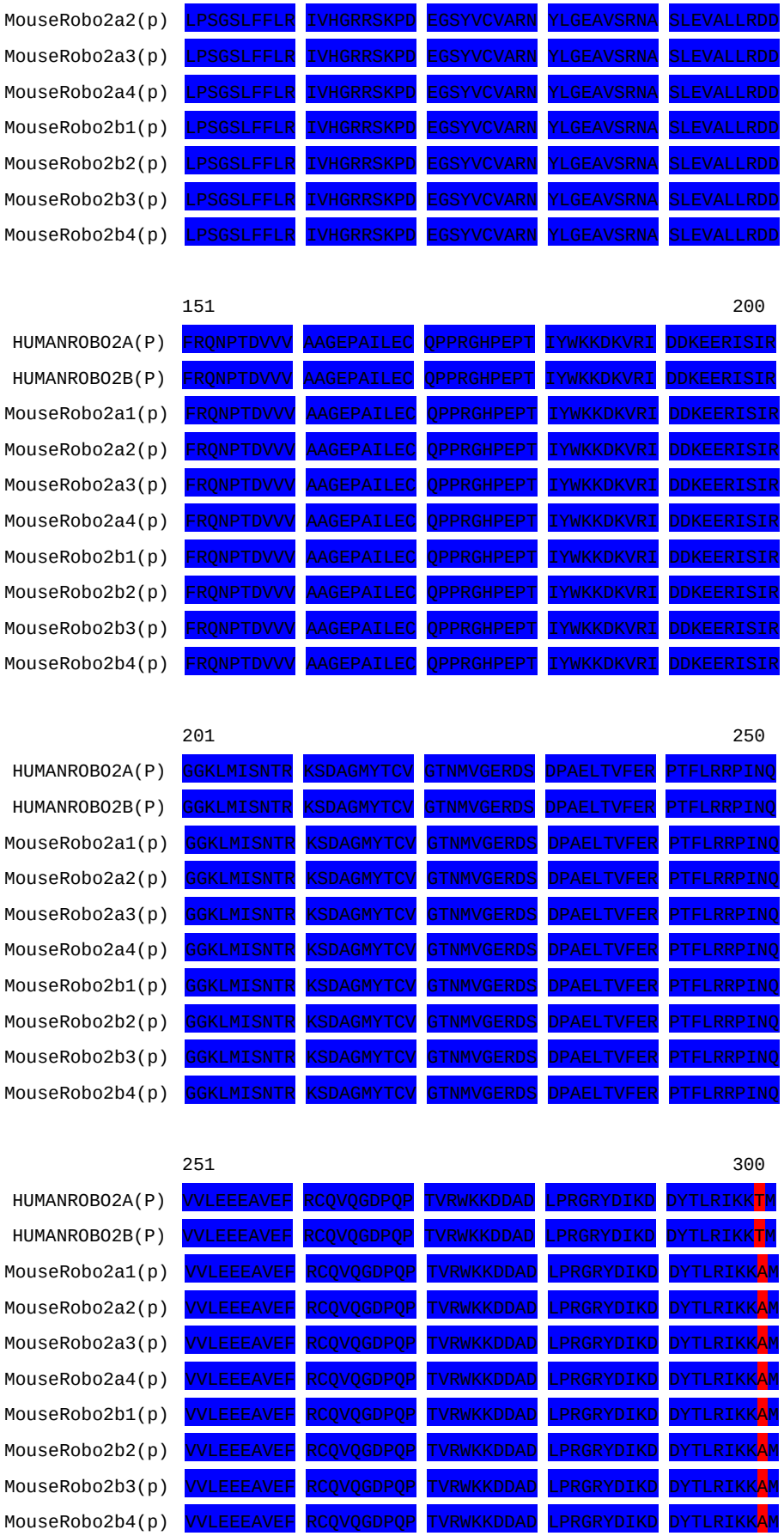
**Supplementary figure 1. Hybridization of BACs on human (HSA) and chimpanzee (PTR) chromosomes.**

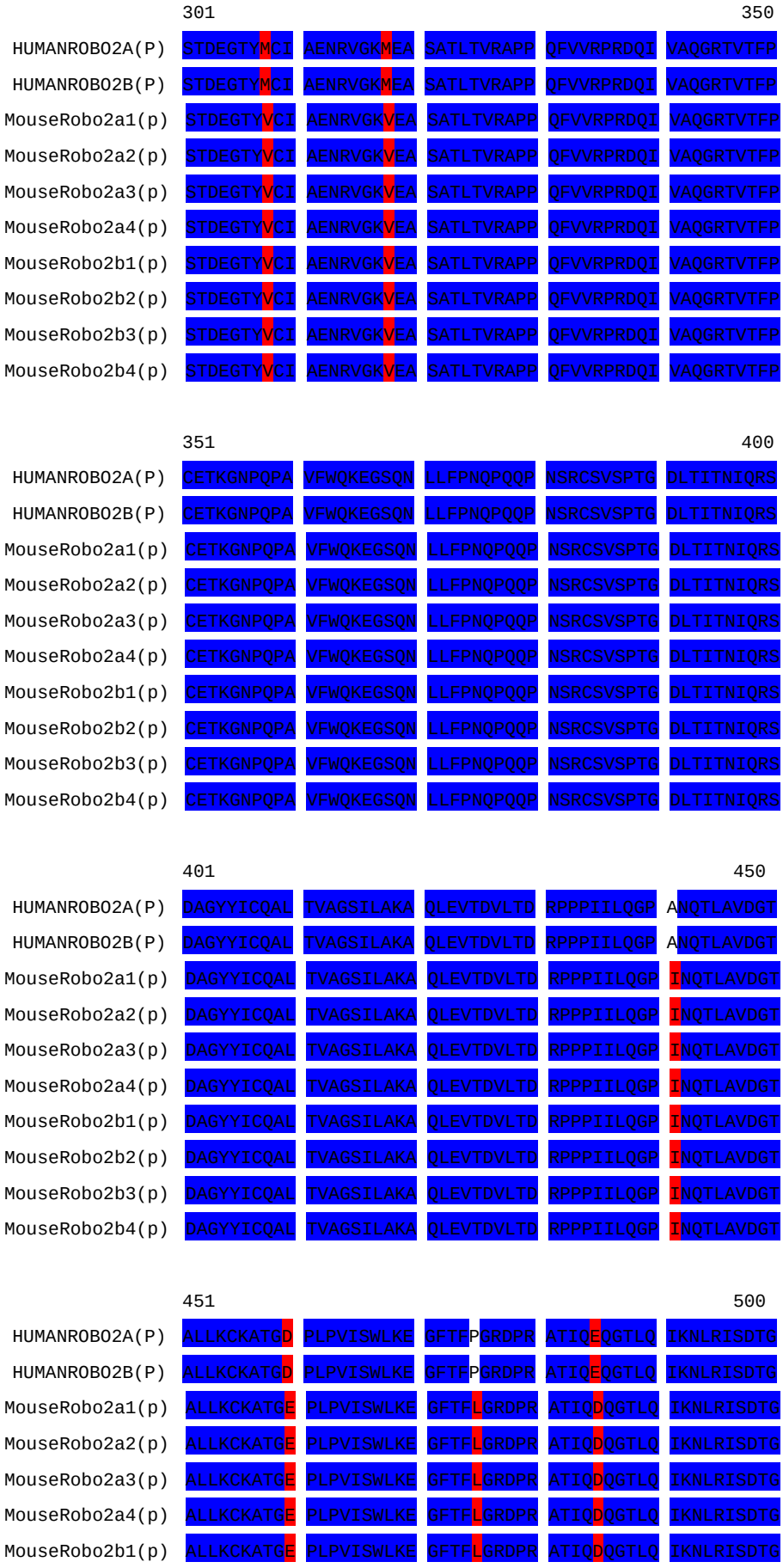


**Supplementary figure 2. Multiple alignment of two human ROBO2 and eight mouse Robo2 variants.**

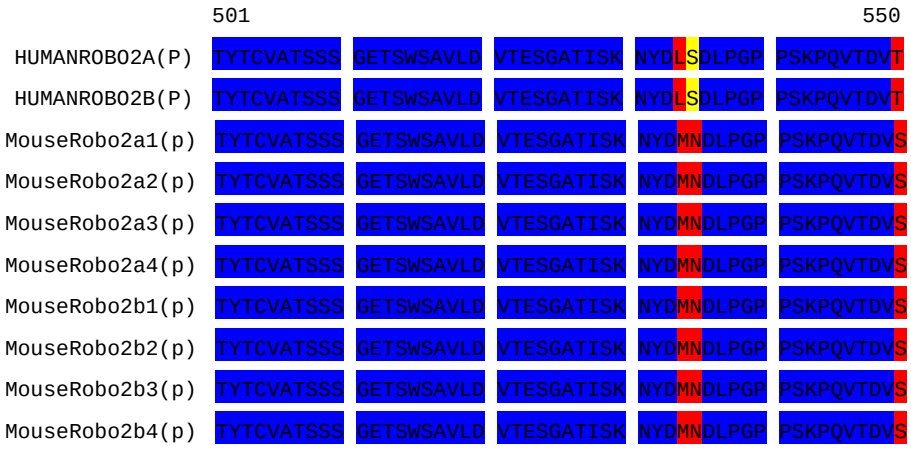
| | | | | | |
|---|---|---|---|---|---|
| 1 HumanROBO2a(p) | from: | 1 | to | 1394 |
| 2 HumanROBO2b(p) | from: | 1 | to | 1378 |
| 3 MouseRobo2a1(p) | from: | 1 | to | 1481 |
| 4 MouseRobo2a2(p) | from: | 1 | to | 1390 |
| 5 MouseRobo2a3(p) | from: | 1 | to | 1523 |
| 6 MouseRobo2a4(p) | from: | 1 | to | 1432 |
| 7 MouseRobo2b1(p) | from: | 1 | to | 1465 |
| 8 MouseRobo2b2(p) | from: | 1 | to | 1375 |
| 9 MouseRobo2b3(p) | from: | 1 | to | 1508 |
| 10 MouseRobo2b4(p) | from: | 1 | to | 1417 |

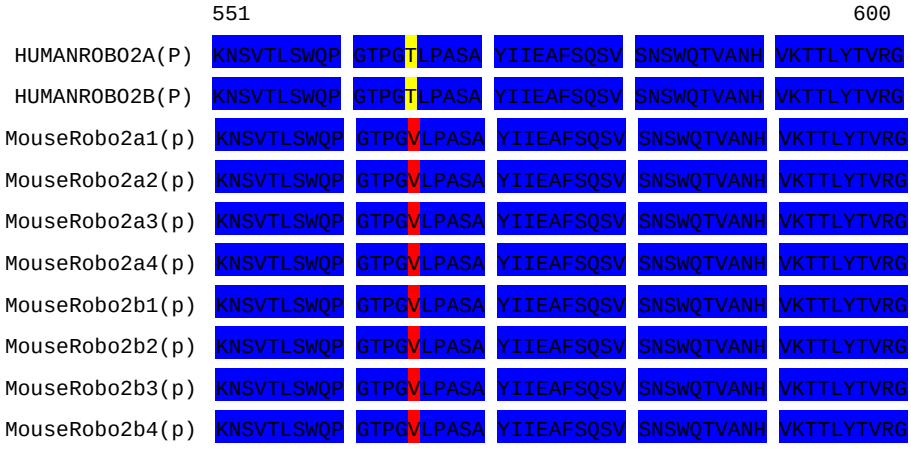


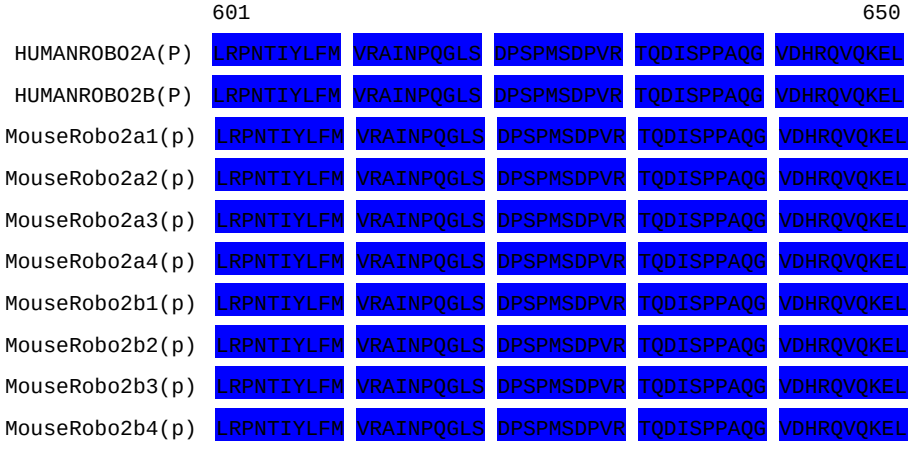

MouseRobo2a2(p) LPSGSLFFLR IVHGRRSKPD EGSYVCVARN YLGEAVSRNA SLEVALLRDD
MouseRobo2a3(p) LPSGSLFFLR IVHGRRSKPD EGSYVCVARN YLGEAVSRNA SLEVALLRDD
MouseRobo2a4(p) LPSGSLFFLR IVHGRRSKPD EGSYVCVARN YLGEAVSRNA SLEVALLRDD
MouseRobo2b1(p) LPSGSLFFLR IVHGRRSKPD EGSYVCVARN YLGEAVSRNA SLEVALLRDD
MouseRobo2b2(p) LPSGSLFFLR IVHGRRSKPD EGSYVCVARN YLGEAVSRNA SLEVALLRDD
MouseRobo2b3(p) LPSGSLFFLR IVHGRRSKPD EGSYVCVARN YLGEAVSRNA SLEVALLRDD
MouseRobo2b4(p) LPSGSLFFLR IVHGRRSKPD EGSYVCVARN YLGEAVSRNA SLEVALLRDD
151 200
HUMANROBO2A(P) FRQNPTDVVV AAGEPAILEC QPPRGHPEPT IYWKKDKVRI DDKEERISIR
HUMANROBO2B(P) FRQNPTDVVV AAGEPAILEC QPPRGHPEPT IYWKKDKVRI DDKEERISIR
MouseRobo2a1(p) FRQNPTDVVV AAGEPAILEC QPPRGHPEPT IYWKKDKVRI DDKEERISIR
MouseRobo2a2(p) FRQNPTDVVV AAGEPAILEC QPPRGHPEPT IYWKKDKVRI DDKEERISIR
MouseRobo2a3(p) FRQNPTDVVV AAGEPAILEC QPPRGHPEPT IYWKKDKVRI DDKEERISIR
MouseRobo2a4(p) FRQNPTDVVV AAGEPAILEC QPPRGHPEPT IYWKKDKVRI DDKEERISIR
MouseRobo2b1(p) FRQNPTDVVV AAGEPAILEC QPPRGHPEPT IYWKKDKVRI DDKEERISIR
MouseRobo2b2(p) FRQNPTDVVV AAGEPAILEC QPPRGHPEPT IYWKKDKVRI DDKEERISIR
MouseRobo2b3(p) FRQNPTDVVV AAGEPAILEC QPPRGHPEPT IYWKKDKVRI DDKEERISIR
MouseRobo2b4(p) FRQNPTDVVV AAGEPAILEC QPPRGHPEPT IYWKKDKVRI DDKEERISIR
201 250
HUMANROBO2A(P) GGKLMISNTR KSDAGMYTCV GTNMVGERDS DPAELTVFER PTFLRRPINQ
HUMANROBO2B(P) GGKLMISNTR KSDAGMYTCV GTNMVGERDS DPAELTVFER PTFLRRPINQ
MouseRobo2a1(p) GGKLMISNTR KSDAGMYTCV GTNMVGERDS DPAELTVFER PTFLRRPINQ
MouseRobo2a2(p) GGKLMISNTR KSDAGMYTCV GTNMVGERDS DPAELTVFER PTFLRRPINQ
MouseRobo2a3(p) GGKLMISNTR KSDAGMYTCV GTNMVGERDS DPAELTVFER PTFLRRPINQ
MouseRobo2a4(p) GGKLMISNTR KSDAGMYTCV GTNMVGERDS DPAELTVFER PTFLRRPINQ
MouseRobo2b1(p) GGKLMISNTR KSDAGMYTCV GTNMVGERDS DPAELTVFER PTFLRRPINQ
MouseRobo2b2(p) GGKLMISNTR KSDAGMYTCV GTNMVGERDS DPAELTVFER PTFLRRPINQ
MouseRobo2b3(p) GGKLMISNTR KSDAGMYTCV GTNMVGERDS DPAELTVFER PTFLRRPINQ
MouseRobo2b4(p) GGKLMISNTR KSDAGMYTCV GTNMVGERDS DPAELTVFER PTFLRRPINQ
251 300
HUMANROBO2A(P) VVLEEEAVEF RCQVQGDPQP TVRWKKDDAD LPRGRYDIKD DYTLRIKKTM
HUMANROBO2B(P) VVLEEEAVEF RCQVQGDPQP TVRWKKDDAD LPRGRYDIKD DYTLRIKKTM
MouseRobo2a1(p) VVLEEEAVEF RCQVQGDPQP TVRWKKDDAD LPRGRYDIKD DYTLRIKKAM
MouseRobo2a2(p) VVLEEEAVEF RCQVQGDPQP TVRWKKDDAD LPRGRYDIKD DYTLRIKKAM
MouseRobo2a3(p) VVLEEEAVEF RCQVQGDPQP TVRWKKDDAD LPRGRYDIKD DYTLRIKKAM
MouseRobo2a4(p) VVLEEEAVEF RCQVQGDPQP TVRWKKDDAD LPRGRYDIKD DYTLRIKKAM
MouseRobo2b1(p) VVLEEEAVEF RCQVQGDPQP TVRWKKDDAD LPRGRYDIKD DYTLRIKKAM
MouseRobo2b2(p) VVLEEEAVEF RCQVQGDPQP TVRWKKDDAD LPRGRYDIKD DYTLRIKKAM
MouseRobo2b3(p) VVLEEEAVEF RCQVQGDPQP TVRWKKDDAD LPRGRYDIKD DYTLRIKKAM
MouseRobo2b4(p) VVLEEEAVEF RCQVQGDPQP TVRWKKDDAD LPRGRYDIKD DYTLRIKKAM


301 350
HUMANROBO2A(P) STDEGTYMCI AENRVGKMEA SATLTVRAPP QFVVRPRDQI VAQGRTVTFP
HUMANROBO2B(P) STDEGTYMCI AENRVGKMEA SATLTVRAPP QFVVRPRDQI VAQGRTVTFP
MouseRobo2a1(p) STDEGTYVCI AENRVGKVEA SATLTVRAPP QFVVRPRDQI VAQGRTVTFP
MouseRobo2a2(p) STDEGTYVCI AENRVGKVEA SATLTVRAPP QFVVRPRDQI VAQGRTVTFP
MouseRobo2a3(p) STDEGTYVCI AENRVGKVEA SATLTVRAPP QFVVRPRDQI VAQGRTVTFP
MouseRobo2a4(p) STDEGTYVCI AENRVGKVEA SATLTVRAPP QFVVRPRDQI VAQGRTVTFP
MouseRobo2b1(p) STDEGTYVCI AENRVGKVEA SATLTVRAPP QFVVRPRDQI VAQGRTVTFP
MouseRobo2b2(p) STDEGTYVCI AENRVGKVEA SATLTVRAPP QFVVRPRDQI VAQGRTVTFP
MouseRobo2b3(p) STDEGTYVCI AENRVGKVEA SATLTVRAPP QFVVRPRDQI VAQGRTVTFP
MouseRobo2b4(p) STDEGTYVCI AENRVGKVEA SATLTVRAPP QFVVRPRDQI VAQGRTVTFP
351 400
HUMANROBO2A(P) CETKGNPQPA VFWQKEGSQN LLFPNQPQQP NSRCSVSPTG DLTITNIQRS
HUMANROBO2B(P) CETKGNPQPA VFWQKEGSQN LLFPNQPQQP NSRCSVSPTG DLTITNIQRS
MouseRobo2a1(p) CETKGNPQPA VFWQKEGSQN LLFPNQPQQP NSRCSVSPTG DLTITNIQRS
MouseRobo2a2(p) CETKGNPQPA VFWQKEGSQN LLFPNQPQQP NSRCSVSPTG DLTITNIQRS
MouseRobo2a3(p) CETKGNPQPA VFWQKEGSQN LLFPNQPQQP NSRCSVSPTG DLTITNIQRS
MouseRobo2a4(p) CETKGNPQPA VFWQKEGSQN LLFPNQPQQP NSRCSVSPTG DLTITNIQRS
MouseRobo2b1(p) CETKGNPQPA VFWQKEGSQN LLFPNQPQQP NSRCSVSPTG DLTITNIQRS
MouseRobo2b2(p) CETKGNPQPA VFWQKEGSQN LLFPNQPQQP NSRCSVSPTG DLTITNIQRS
MouseRobo2b3(p) CETKGNPQPA VFWQKEGSQN LLFPNQPQQP NSRCSVSPTG DLTITNIQRS
MouseRobo2b4(p) CETKGNPQPA VFWQKEGSQN LLFPNQPQQP NSRCSVSPTG DLTITNIQRS
401 450
HUMANROBO2A(P) DAGYYICQAL TVAGSILAKA QLEVTDVLTD RPPPIILQGP ANQTLAVDGT
HUMANROBO2B(P) DAGYYICQAL TVAGSILAKA QLEVTDVLTD RPPPIILQGP ANQTLAVDGT
MouseRobo2a1(p) DAGYYICQAL TVAGSILAKA QLEVTDVLTD RPPPIILQGP INQTLAVDGT
MouseRobo2a2(p) DAGYYICQAL TVAGSILAKA QLEVTDVLTD RPPPIILQGP INQTLAVDGT
MouseRobo2a3(p) DAGYYICQAL TVAGSILAKA QLEVTDVLTD RPPPIILQGP INQTLAVDGT
MouseRobo2a4(p) DAGYYICQAL TVAGSILAKA QLEVTDVLTD RPPPIILQGP INQTLAVDGT
MouseRobo2b1(p) DAGYYICQAL TVAGSILAKA QLEVTDVLTD RPPPIILQGP INQTLAVDGT
MouseRobo2b2(p) DAGYYICQAL TVAGSILAKA QLEVTDVLTD RPPPIILQGP INQTLAVDGT
MouseRobo2b3(p) DAGYYICQAL TVAGSILAKA QLEVTDVLTD RPPPIILQGP INQTLAVDGT
MouseRobo2b4(p) DAGYYICQAL TVAGSILAKA QLEVTDVLTD RPPPIILQGP INQTLAVDGT
451 500
HUMANROBO2A(P) ALLKCKATGD PLPVISWLKE GFTPPGRDPR ATIQEQGTLQ IKNLRISDTG
HUMANROBO2B(P) ALLKCKATGD PLPVISWLKE GFTPPGRDPR ATIQEQGTLQ IKNLRISDTG
MouseRobo2a1(p) ALLKCKATGE PLPVISWLKE GFTFLGRDPR ATIQDQGTLQ IKNLRISDTG
MouseRobo2a2(p) ALLKCKATGE PLPVISWLKE GFTFLGRDPR ATIQDQGTLQ IKNLRISDTG
MouseRobo2a3(p) ALLKCKATGE PLPVISWLKE GFTFLGRDPR ATIQDQGTLQ IKNLRISDTG
MouseRobo2a4(p) ALLKCKATGE PLPVISWLKE GFTFLGRDPR ATIQDQGTLQ IKNLRISDTG
MouseRobo2b1(p) ALLKCKATGE PLPVISWLKE GFTFLGRDPR ATIQDQGTLQ IKNLRISDTG

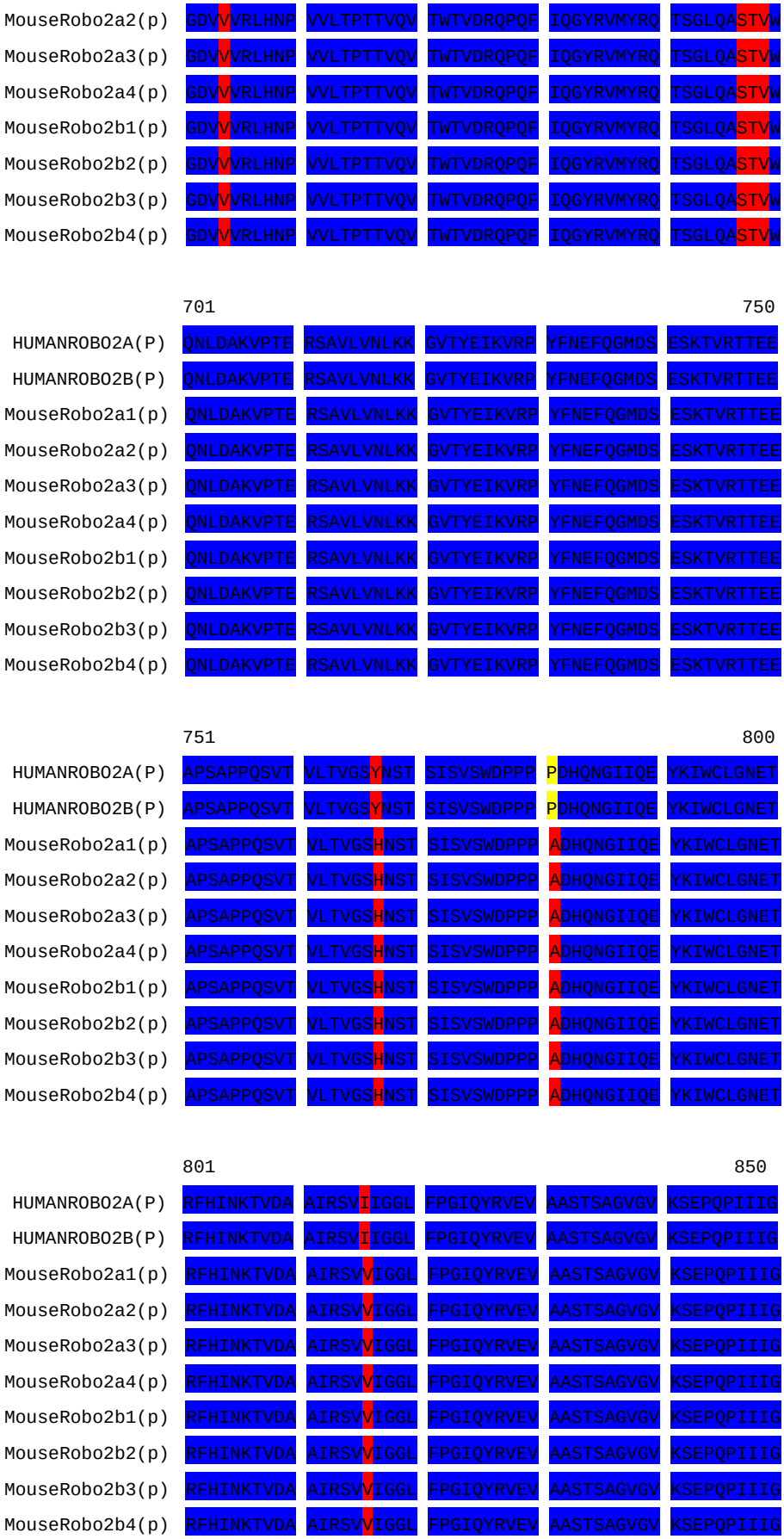

MouseRobo2b2(p) ALLKCKATGE PLPVISWLKE GFTFLGRDPR ATIQDQGTLQ IKNLRISDTG
MouseRobo2b3(p) ALLKCKATGE PLPVISWLKE GFTFLGRDPR ATIQDQGTLQ IKNLRISDTG
MouseRobo2b4(p) ALLKCKATGE PLPVISWLKE GFTFLGRDPR ATIQDQGTLQ IKNLRISDTG

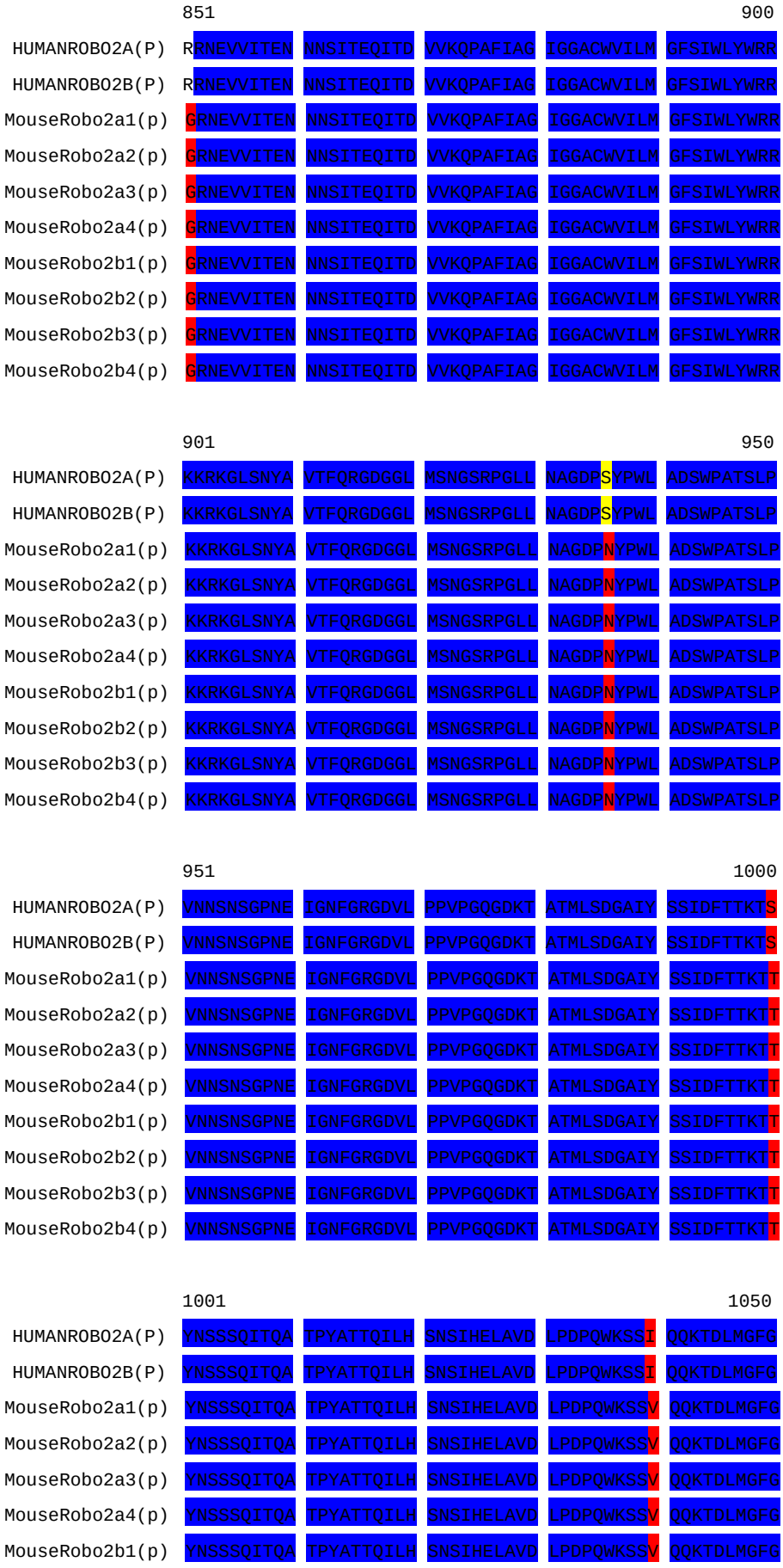

501 550
HUMANROBO2A(P) TYTCVATSSS GETSWSAVLD VTESGATISK NYDLSDLPGP PSKPQVTDVT
HUMANROBO2B(P) TYTCVATSSS GETSWSAVLD VTESGATISK NYDLSDLPGP PSKPQVTDVT
MouseRobo2a1(p) TYTCVATSSS GETSWSAVLD VTESGATISK NYDMNDLPGP PSKPQVTDVS
MouseRobo2a2(p) TYTCVATSSS GETSWSAVLD VTESGATISK NYDMNDLPGP PSKPQVTDVS
MouseRobo2a3(p) TYTCVATSSS GETSWSAVLD VTESGATISK NYDMNDLPGP PSKPQVTDVS
MouseRobo2a4(p) TYTCVATSSS GETSWSAVLD VTESGATISK NYDMNDLPGP PSKPQVTDVS
MouseRobo2b1(p) TYTCVATSSS GETSWSAVLD VTESGATISK NYDMNDLPGP PSKPQVTDVS
MouseRobo2b2(p) TYTCVATSSS GETSWSAVLD VTESGATISK NYDMNDLPGP PSKPQVTDVS
MouseRobo2b3(p) TYTCVATSSS GETSWSAVLD VTESGATISK NYDMNDLPGP PSKPQVTDVS
MouseRobo2b4(p) TYTCVATSSS GETSWSAVLD VTESGATISK NYDMNDLPGP PSKPQVTDVS

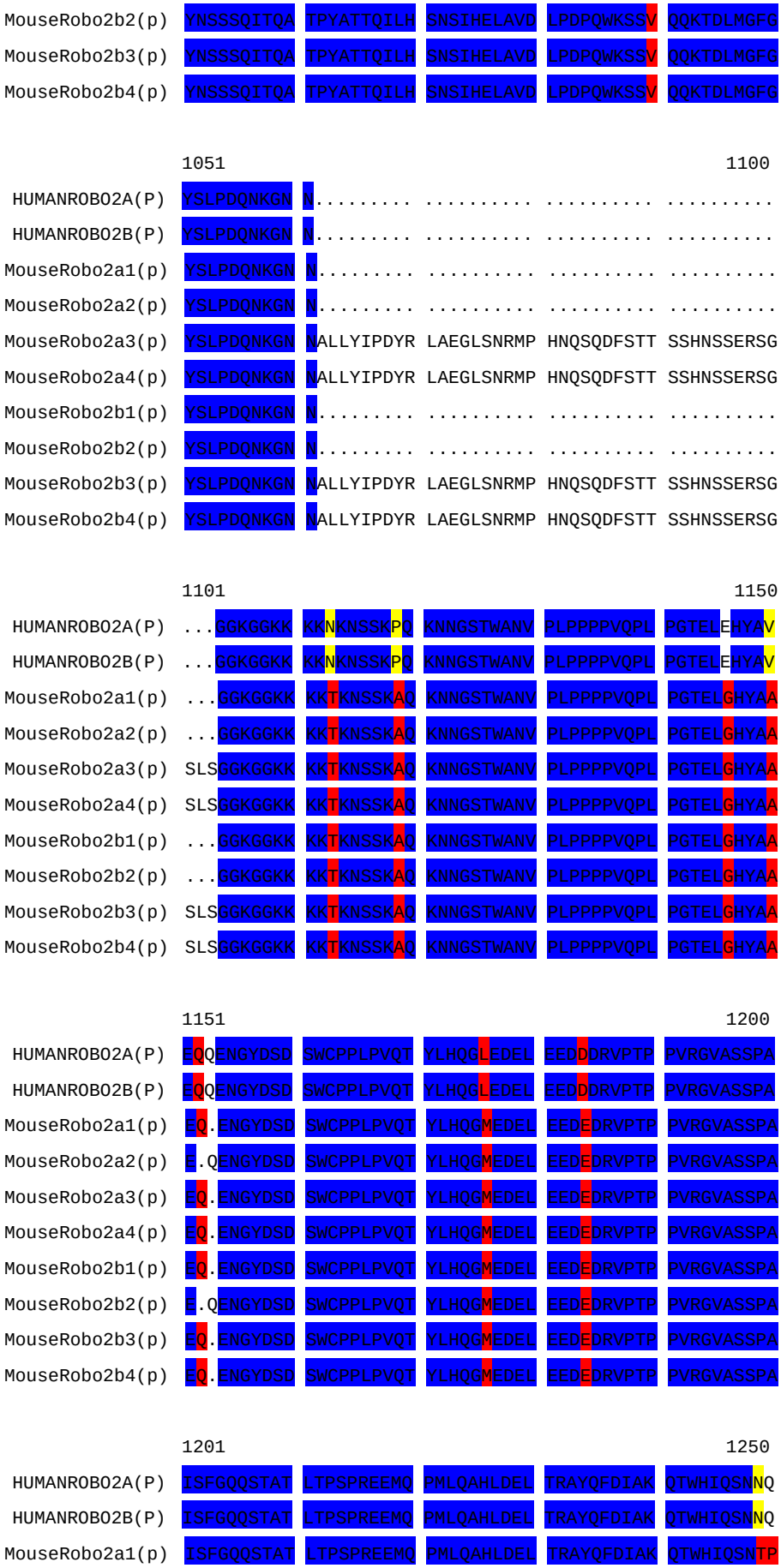

551 600
HUMANROBO2A(P) KNSVTLSWQP GTPGTLPASA YIIEAFSQSV SNSWQTVANH VKTTLYTVRG
HUMANROBO2B(P) KNSVTLSWQP GTPGTLPASA YIIEAFSQSV SNSWQTVANH VKTTLYTVRG
MouseRobo2a1(p) KNSVTLSWQP GTPGVLPASA YIIEAFSQSV SNSWQTVANH VKTTLYTVRG
MouseRobo2a2(p) KNSVTLSWQP GTPGVLPASA YIIEAFSQSV SNSWQTVANH VKTTLYTVRG
MouseRobo2a3(p) KNSVTLSWQP GTPGVLPASA YIIEAFSQSV SNSWQTVANH VKTTLYTVRG
MouseRobo2a4(p) KNSVTLSWQP GTPGVLPASA YIIEAFSQSV SNSWQTVANH VKTTLYTVRG
MouseRobo2b1(p) KNSVTLSWQP GTPGVLPASA YIIEAFSQSV SNSWQTVANH VKTTLYTVRG
MouseRobo2b2(p) KNSVTLSWQP GTPGVLPASA YIIEAFSQSV SNSWQTVANH VKTTLYTVRG
MouseRobo2b3(p) KNSVTLSWQP GTPGVLPASA YIIEAFSQSV SNSWQTVANH VKTTLYTVRG
MouseRobo2b4(p) KNSVTLSWQP GTPGVLPASA YIIEAFSQSV SNSWQTVANH VKTTLYTVRG

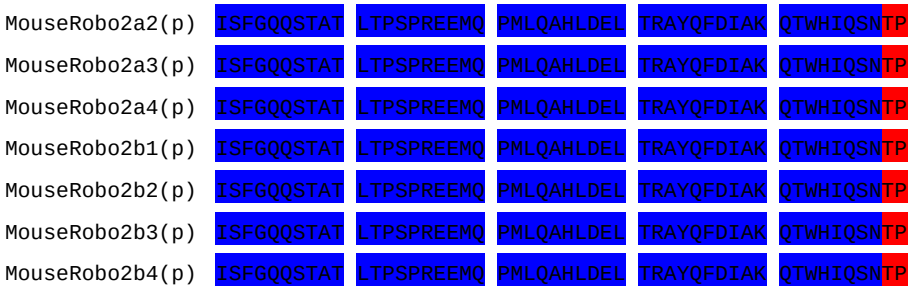

601 650
HUMANROBO2A(P) LRPNTIYLFM VRAINPQGLS DPSPMSDPVR TQDISPPAQG VDHRQVQKEL
HUMANROBO2B(P) LRPNTIYLFM VRAINPQGLS DPSPMSDPVR TQDISPPAQG VDHRQVQKEL
MouseRobo2a1(p) LRPNTIYLFM VRAINPQGLS DPSPMSDPVR TQDISPPAQG VDHRQVQKEL
MouseRobo2a2(p) LRPNTIYLFM VRAINPQGLS DPSPMSDPVR TQDISPPAQG VDHRQVQKEL
MouseRobo2a3(p) LRPNTIYLFM VRAINPQGLS DPSPMSDPVR TQDISPPAQG VDHRQVQKEL
MouseRobo2a4(p) LRPNTIYLFM VRAINPQGLS DPSPMSDPVR TQDISPPAQG VDHRQVQKEL
MouseRobo2b1(p) LRPNTIYLFM VRAINPQGLS DPSPMSDPVR TQDISPPAQG VDHRQVQKEL
MouseRobo2b2(p) LRPNTIYLFM VRAINPQGLS DPSPMSDPVR TQDISPPAQG VDHRQVQKEL
MouseRobo2b3(p) LRPNTIYLFM VRAINPQGLS DPSPMSDPVR TQDISPPAQG VDHRQVQKEL
MouseRobo2b4(p) LRPNTIYLFM VRAINPQGLS DPSPMSDPVR TQDISPPAQG VDHRQVQKEL

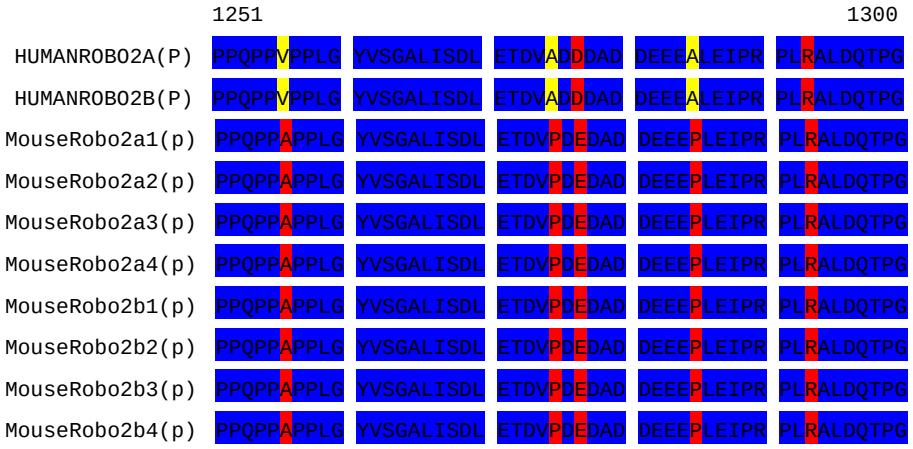

651 700
HUMANROBO2A(P) GDVLVRLHNP VVLTPTTVQV TWTVDRQPQF IQGYRVMYRQ TSGLQATSSW
HUMANROBO2B(P) GDVLVRLHNP VVLTPTTVQV TWTVDRQPQF IQGYRVMYRQ TSGLQATSSW
MouseRobo2a1(p) GDVVVRLHNP VVLTPTTVQV TWTVDRQPQF IQGYRVMYRQ TSGLQASTVW

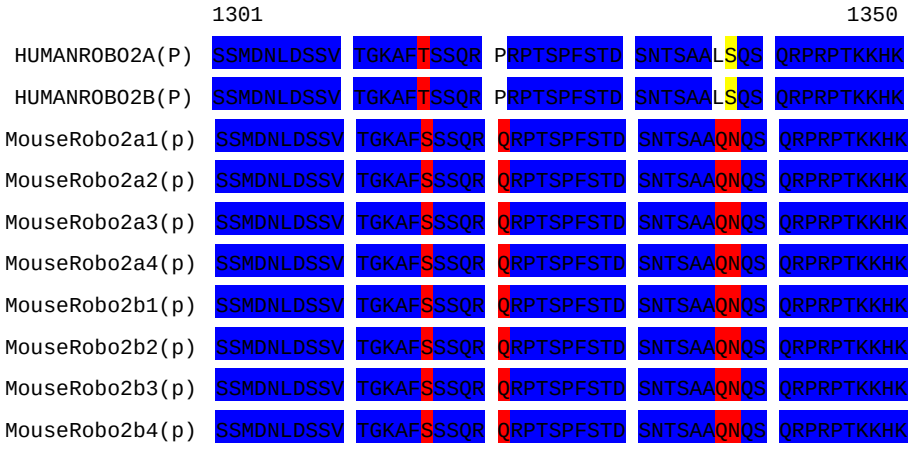

MouseRobo2a2(p) GDVVVRLHNP VVLTPTTVQV TWTVDRQPQF IQGYRVMYRQ TSGLQASTVW
MouseRobo2a3(p) GDVVVRLHNP VVLTPTTVQV TWTVDRQPQF IQGYRVMYRQ TSGLQASTVW
MouseRobo2a4(p) GDVVVRLHNP VVLTPTTVQV TWTVDRQPQF IQGYRVMYRQ TSGLQASTVW
MouseRobo2b1(p) GDVVVRLHNP VVLTPTTVQV TWTVDRQPQF IQGYRVMYRQ TSGLQASTVW
MouseRobo2b2(p) GDVVVRLHNP VVLTPTTVQV TWTVDRQPQF IQGYRVMYRQ TSGLQASTVW
MouseRobo2b3(p) GDVVVRLHNP VVLTPTTVQV TWTVDRQPQF IQGYRVMYRQ TSGLQASTVW
MouseRobo2b4(p) GDVVVRLHNP VVLTPTTVQV TWTVDRQPQF IQGYRVMYRQ TSGLQASTVW
701
750
HUMANROBO2A(P) QNLDAKVPTE RSAVLVNLKK GVTYEIKVRP YFNEFQGMDS ESKTVRTTEE
HUMANROBO2B(P) QNLDAKVPTE RSAVLVNLKK GVTYEIKVRP YFNEFQGMDS ESKTVRTTEE
MouseRobo2a1(p) QNLDAKVPTE RSAVLVNLKK GVTYEIKVRP YFNEFQGMDS ESKTVRTTEE
MouseRobo2a2(p) QNLDAKVPTE RSAVLVNLKK GVTYEIKVRP YFNEFQGMDS ESKTVRTTEE
MouseRobo2a3(p) QNLDAKVPTE RSAVLVNLKK GVTYEIKVRP YFNEFQGMDS ESKTVRTTEE
MouseRobo2a4(p) QNLDAKVPTE RSAVLVNLKK GVTYEIKVRP YFNEFQGMDS ESKTVRTTEE
MouseRobo2b1(p) QNLDAKVPTE RSAVLVNLKK GVTYEIKVRP YFNEFQGMDS ESKTVRTTEE
MouseRobo2b2(p) QNLDAKVPTE RSAVLVNLKK GVTYEIKVRP YFNEFQGMDS ESKTVRTTEE
MouseRobo2b3(p) QNLDAKVPTE RSAVLVNLKK GVTYEIKVRP YFNEFQGMDS ESKTVRTTEE
MouseRobo2b4(p) QNLDAKVPTE RSAVLVNLKK GVTYEIKVRP YFNEFQGMDS ESKTVRTTEE
751
800
HUMANROBO2A(P) APSAPPQSVT VLTVGSYNST SISVSWDPPP PDHQNGIIQE YKIWCLGNET
HUMANROBO2B(P) APSAPPQSVT VLTVGSYNST SISVSWDPPP PDHQNGIIQE YKIWCLGNET
MouseRobo2a1(p) APSAPPQSVT VLTVGSHNST SISVSWDPPP ADHQNGIIQE YKIWCLGNET
MouseRobo2a2(p) APSAPPQSVT VLTVGSHNST SISVSWDPPP ADHQNGIIQE YKIWCLGNET
MouseRobo2a3(p) APSAPPQSVT VLTVGSHNST SISVSWDPPP ADHQNGIIQE YKIWCLGNET
MouseRobo2a4(p) APSAPPQSVT VLTVGSHNST SISVSWDPPP ADHQNGIIQE YKIWCLGNET
MouseRobo2b1(p) APSAPPQSVT VLTVGSHNST SISVSWDPPP ADHQNGIIQE YKIWCLGNET
MouseRobo2b2(p) APSAPPQSVT VLTVGSHNST SISVSWDPPP ADHQNGIIQE YKIWCLGNET
MouseRobo2b3(p) APSAPPQSVT VLTVGSHNST SISVSWDPPP ADHQNGIIQE YKIWCLGNET
MouseRobo2b4(p) APSAPPQSVT VLTVGSHNST SISVSWDPPP ADHQNGIIQE YKIWCLGNET
801
850
HUMANROBO2A(P) RFHINKTVDA AIRSVIIGGL FPGIQYRVEV AASTSAGVGV KSEPQPIIIG
HUMANROBO2B(P) RFHINKTVDA AIRSVIIGGL FPGIQYRVEV AASTSAGVGV KSEPQPIIIG
MouseRobo2a1(p) RFHINKTVDA AIRSVVIGGL FPGIQYRVEV AASTSAGVGV KSEPQPIIIG
MouseRobo2a2(p) RFHINKTVDA AIRSVVIGGL FPGIQYRVEV AASTSAGVGV KSEPQPIIIG
MouseRobo2a3(p) RFHINKTVDA AIRSVVIGGL FPGIQYRVEV AASTSAGVGV KSEPQPIIIG
MouseRobo2a4(p) RFHINKTVDA AIRSVVIGGL FPGIQYRVEV AASTSAGVGV KSEPQPIIIG
MouseRobo2b1(p) RFHINKTVDA AIRSVVIGGL FPGIQYRVEV AASTSAGVGV KSEPQPIIIG
MouseRobo2b2(p) RFHINKTVDA AIRSVVIGGL FPGIQYRVEV AASTSAGVGV KSEPQPIIIG
MouseRobo2b3(p) RFHINKTVDA AIRSVVIGGL FPGIQYRVEV AASTSAGVGV KSEPQPIIIG
MouseRobo2b4(p) RFHINKTVDA AIRSVVIGGL FPGIQYRVEV AASTSAGVGV KSEPQPIIIG

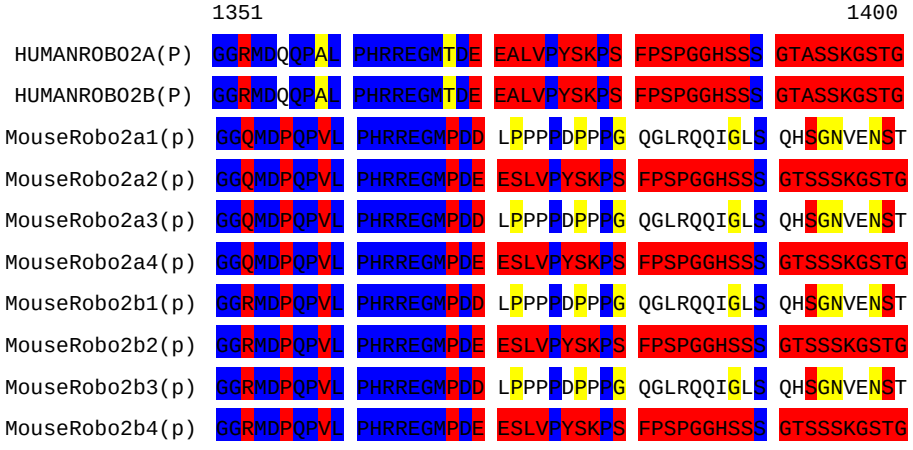

851
900
HUMANROBO2A(P) RRNEVVITEN NNSITEQITD VVKQPAFIAG IGGACWVILM GFSIWLYWRR
HUMANROBO2B(P) RRNEVVITEN NNSITEQITD VVKQPAFIAG IGGACWVILM GFSIWLYWRR
MouseRobo2a1(p) GRNEVVITEN NNSITEQITD VVKQPAFIAG IGGACWVILM GFSIWLYWRR
MouseRobo2a2(p) GRNEVVITEN NNSITEQITD VVKQPAFIAG IGGACWVILM GFSIWLYWRR
MouseRobo2a3(p) GRNEVVITEN NNSITEQITD VVKQPAFIAG IGGACWVILM GFSIWLYWRR
MouseRobo2a4(p) GRNEVVITEN NNSITEQITD VVKQPAFIAG IGGACWVILM GFSIWLYWRR
MouseRobo2b1(p) GRNEVVITEN NNSITEQITD VVKQPAFIAG IGGACWVILM GFSIWLYWRR
MouseRobo2b2(p) GRNEVVITEN NNSITEQITD VVKQPAFIAG IGGACWVILM GFSIWLYWRR
MouseRobo2b3(p) GRNEVVITEN NNSITEQITD VVKQPAFIAG IGGACWVILM GFSIWLYWRR
MouseRobo2b4(p) GRNEVVITEN NNSITEQITD VVKQPAFIAG IGGACWVILM GFSIWLYWRR
901
950
HUMANROBO2A(P) KKRKGLSNYA VTFQRGDGGL MSNGSRPGLL NAGDPSYPWL ADSWPATSLP
HUMANROBO2B(P) KKRKGLSNYA VTFQRGDGGL MSNGSRPGLL NAGDPSYPWL ADSWPATSLP
MouseRobo2a1(p) KKRKGLSNYA VTFQRGDGGL MSNGSRPGLL NAGDPNYPWL ADSWPATSLP
MouseRobo2a2(p) KKRKGLSNYA VTFQRGDGGL MSNGSRPGLL NAGDPNYPWL ADSWPATSLP
MouseRobo2a3(p) KKRKGLSNYA VTFQRGDGGL MSNGSRPGLL NAGDPNYPWL ADSWPATSLP
MouseRobo2a4(p) KKRKGLSNYA VTFQRGDGGL MSNGSRPGLL NAGDPNYPWL ADSWPATSLP
MouseRobo2b1(p) KKRKGLSNYA VTFQRGDGGL MSNGSRPGLL NAGDPNYPWL ADSWPATSLP
MouseRobo2b2(p) KKRKGLSNYA VTFQRGDGGL MSNGSRPGLL NAGDPNYPWL ADSWPATSLP
MouseRobo2b3(p) KKRKGLSNYA VTFQRGDGGL MSNGSRPGLL NAGDPNYPWL ADSWPATSLP
MouseRobo2b4(p) KKRKGLSNYA VTFQRGDGGL MSNGSRPGLL NAGDPNYPWL ADSWPATSLP
951
1000
HUMANROBO2A(P) VNNSNSGPNE IGNFGRGDVL PPVPGQGDKT ATMLSDGAIY SSIDFTTKTS
HUMANROBO2B(P) VNNSNSGPNE IGNFGRGDVL PPVPGQGDKT ATMLSDGAIY SSIDFTTKTS
MouseRobo2a1(p) VNNSNSGPNE IGNFGRGDVL PPVPGQGDKT ATMLSDGAIY SSIDFTTKTT
MouseRobo2a2(p) VNNSNSGPNE IGNFGRGDVL PPVPGQGDKT ATMLSDGAIY SSIDFTTKTT
MouseRobo2a3(p) VNNSNSGPNE IGNFGRGDVL PPVPGQGDKT ATMLSDGAIY SSIDFTTKTT
MouseRobo2a4(p) VNNSNSGPNE IGNFGRGDVL PPVPGQGDKT ATMLSDGAIY SSIDFTTKTT
MouseRobo2b1(p) VNNSNSGPNE IGNFGRGDVL PPVPGQGDKT ATMLSDGAIY SSIDFTTKTT
MouseRobo2b2(p) VNNSNSGPNE IGNFGRGDVL PPVPGQGDKT ATMLSDGAIY SSIDFTTKTT
MouseRobo2b3(p) VNNSNSGPNE IGNFGRGDVL PPVPGQGDKT ATMLSDGAIY SSIDFTTKTT
MouseRobo2b4(p) VNNSNSGPNE IGNFGRGDVL PPVPGQGDKT ATMLSDGAIY SSIDFTTKTT
1001
1050
HUMANROBO2A(P) YNSSSQITQA TPYATTQILH SNSIHELAVD LPDPQWKSSI QQKTDLMGFG
HUMANROBO2B(P) YNSSSQITQA TPYATTQILH SNSIHELAVD LPDPQWKSSI QQKTDLMGFG
MouseRobo2a1(p) YNSSSQITQA TPYATTQILH SNSIHELAVD LPDPQWKSSV QQKTDLMGFG
MouseRobo2a2(p) YNSSSQITQA TPYATTQILH SNSIHELAVD LPDPQWKSSV QQKTDLMGFG
MouseRobo2a3(p) YNSSSQITQA TPYATTQILH SNSIHELAVD LPDPQWKSSV QQKTDLMGFG
MouseRobo2a4(p) YNSSSQITQA TPYATTQILH SNSIHELAVD LPDPQWKSSV QQKTDLMGFG
MouseRobo2b1(p) YNSSSQITQA TPYATTQILH SNSIHELAVD LPDPQWKSSV QQKTDLMGFG

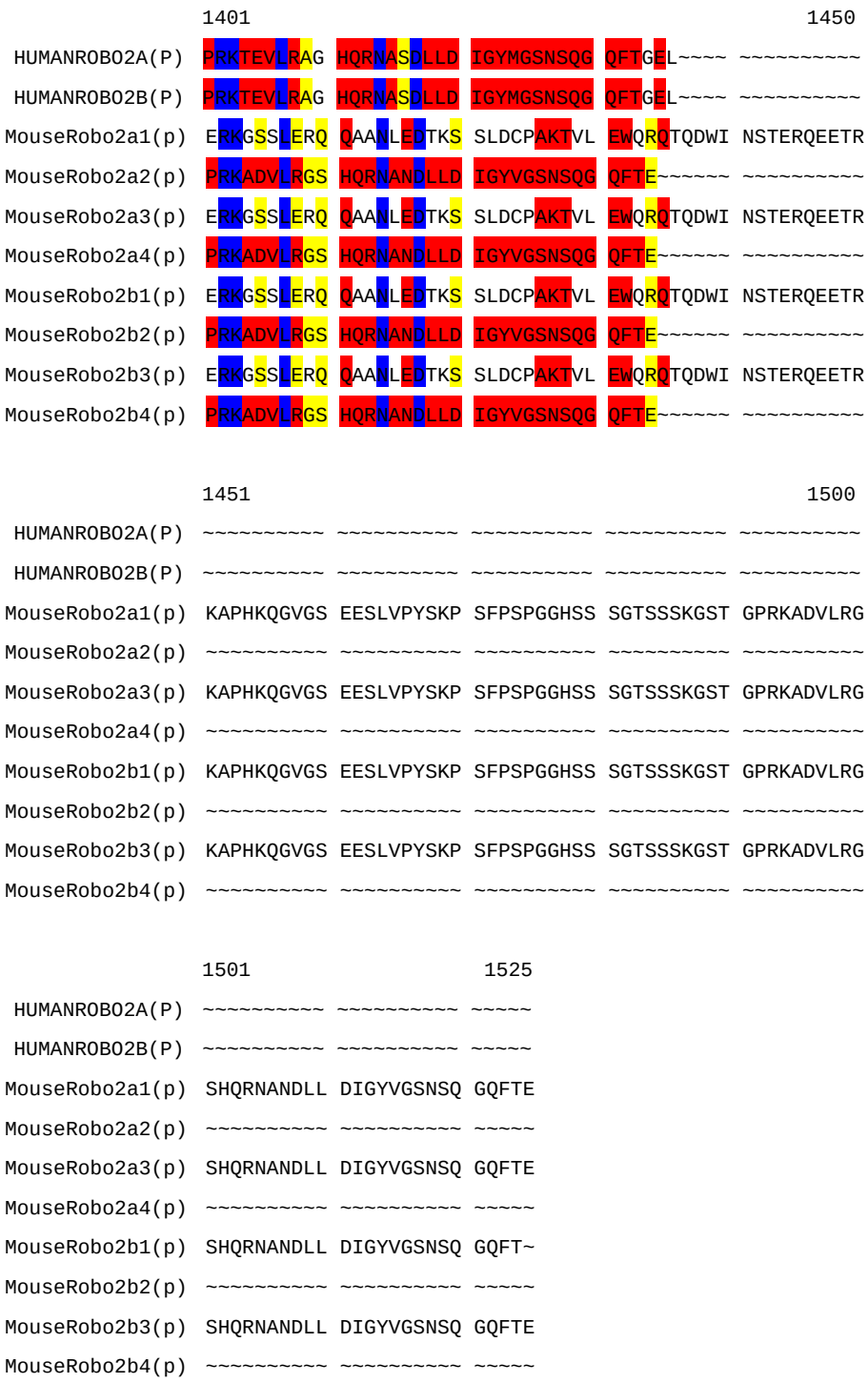

MouseRobo2b2(p) YNSSSQITQA TPYATTQILH SNSIHELAVD LPDPQWKSSV QQKTDLMGFG
MouseRobo2b3(p) YNSSSQITQA TPYATTQILH SNSIHELAVD LPDPQWKSSV QQKTDLMGFG
MouseRobo2b4(p) YNSSSQITQA TPYATTQILH SNSIHELAVD LPDPQWKSSV QQKTDLMGFG
1051
1100
HUMANROBO2A(P) YSLPDQNKGN N..... ..... ..... .....
HUMANROBO2B(P) YSLPDQNKGN N..... ..... ..... .....
MouseRobo2a1(p) YSLPDQNKGN N..... ..... ..... .....
MouseRobo2a2(p) YSLPDQNKGN N..... ..... ..... .....
MouseRobo2a3(p) YSLPDQNKGN NALLYIPDYR LAEGLSNRMP HNQSQDFSTT SSHNSSERSG
MouseRobo2a4(p) YSLPDQNKGN NALLYIPDYR LAEGLSNRMP HNQSQDFSTT SSHNSSERSG
MouseRobo2b1(p) YSLPDQNKGN N..... ..... ..... .....
MouseRobo2b2(p) YSLPDQNKGN N..... ..... ..... .....
MouseRobo2b3(p) YSLPDQNKGN NALLYIPDYR LAEGLSNRMP HNQSQDFSTT SSHNSSERSG
MouseRobo2b4(p) YSLPDQNKGN NALLYIPDYR LAEGLSNRMP HNQSQDFSTT SSHNSSERSG
1101
1150
HUMANROBO2A(P) ...GGKGGKK KKNKNSSKPQ KNNGSTWANV PLPPPPVQPL PGTELEHYAV
HUMANROBO2B(P) ...GGKGGKK KKNKNSSKPQ KNNGSTWANV PLPPPPVQPL PGTELEHYAV
MouseRobo2a1(p) ...GGKGGKK KKTKNSSKAQ KNNGSTWANV PLPPPPVQPL PGTELGHYAA
MouseRobo2a2(p) ...GGKGGKK KKTKNSSKAQ KNNGSTWANV PLPPPPVQPL PGTELGHYAA
MouseRobo2a3(p) SLSGGKGGKK KKTKNSSKAQ KNNGSTWANV PLPPPPVQPL PGTELGHYAA
MouseRobo2a4(p) SLSGGKGGKK KKTKNSSKAQ KNNGSTWANV PLPPPPVQPL PGTELGHYAA
MouseRobo2b1(p) ...GGKGGKK KKTKNSSKAQ KNNGSTWANV PLPPPPVQPL PGTELGHYAA
MouseRobo2b2(p) ...GGKGGKK KKTKNSSKAQ KNNGSTWANV PLPPPPVQPL PGTELGHYAA
MouseRobo2b3(p) SLSGGKGGKK KKTKNSSKAQ KNNGSTWANV PLPPPPVQPL PGTELGHYAA
MouseRobo2b4(p) SLSGGKGGKK KKTKNSSKAQ KNNGSTWANV PLPPPPVQPL PGTELGHYAA
1151
1200
HUMANROBO2A(P) EQQENGYDSD SWCPPLPVQT YLHQGLEDEL EEDDDRVPTP PVRGVASSPA
HUMANROBO2B(P) EQQENGYDSD SWCPPLPVQT YLHQGLEDEL EEDDDRVPTP PVRGVASSPA
MouseRobo2a1(p) EQ.ENGYDSD SWCPPLPVQT YLHQGMEDEL EEDEDRVPTP PVRGVASSPA
MouseRobo2a2(p) E.QENGYDSD SWCPPLPVQT YLHQGMEDEL EEDEDRVPTP PVRGVASSPA
MouseRobo2a3(p) EQ.ENGYDSD SWCPPLPVQT YLHQGMEDEL EEDEDRVPTP PVRGVASSPA
MouseRobo2a4(p) EQ.ENGYDSD SWCPPLPVQT YLHQGMEDEL EEDEDRVPTP PVRGVASSPA
MouseRobo2b1(p) EQ.ENGYDSD SWCPPLPVQT YLHQGMEDEL EEDEDRVPTP PVRGVASSPA
MouseRobo2b2(p) E.QENGYDSD SWCPPLPVQT YLHQGMEDEL EEDEDRVPTP PVRGVASSPA
MouseRobo2b3(p) EQ.ENGYDSD SWCPPLPVQT YLHQGMEDEL EEDEDRVPTP PVRGVASSPA
MouseRobo2b4(p) EQ.ENGYDSD SWCPPLPVQT YLHQGMEDEL EEDEDRVPTP PVRGVASSPA
1201
1250
HUMANROBO2A(P) ISFGQQSTAT LTPSPREEMQ PMLQAHLDEL TRAYQFDIAK QTWHIQSNNQ
HUMANROBO2B(P) ISFGQQSTAT LTPSPREEMQ PMLQAHLDEL TRAYQFDIAK QTWHIQSNNQ
MouseRobo2a1(p) ISFGQQSTAT LTPSPREEMQ PMLQAHLDEL TRAYQFDIAK QTWHIQSNTP


MouseRobo2a2(p) ISFGQQSTAT LTPSPREEMQ PMLQAHLDEL TRAYQFDIAK QTWHIQSNTP
MouseRobo2a3(p) ISFGQQSTAT LTPSPREEMQ PMLQAHLDEL TRAYQFDIAK QTWHIQSNTP
MouseRobo2a4(p) ISFGQQSTAT LTPSPREEMQ PMLQAHLDEL TRAYQFDIAK QTWHIQSNTP
MouseRobo2b1(p) ISFGQQSTAT LTPSPREEMQ PMLQAHLDEL TRAYQFDIAK QTWHIQSNTP
MouseRobo2b2(p) ISFGQQSTAT LTPSPREEMQ PMLQAHLDEL TRAYQFDIAK QTWHIQSNTP
MouseRobo2b3(p) ISFGQQSTAT LTPSPREEMQ PMLQAHLDEL TRAYQFDIAK QTWHIQSNTP
MouseRobo2b4(p) ISFGQQSTAT LTPSPREEMQ PMLQAHLDEL TRAYQFDIAK QTWHIQSNTP


1251
1300
HUMANROBO2A(P) PPQPPVPPLG YVSGALISDL ETDVADDDAD DEEEALEIPR PLRALDQTPG
HUMANROBO2B(P) PPQPPVPPLG YVSGALISDL ETDVADDDAD DEEEALEIPR PLRALDQTPG
MouseRobo2a1(p) PPQPPAPPLG YVSGALISDL ETDVPDEDAD DEEEPLEIPR PLRALDQTPG
MouseRobo2a2(p) PPQPPAPPLG YVSGALISDL ETDVPDEDAD DEEEPLEIPR PLRALDQTPG
MouseRobo2a3(p) PPQPPAPPLG YVSGALISDL ETDVPDEDAD DEEEPLEIPR PLRALDQTPG
MouseRobo2a4(p) PPQPPAPPLG YVSGALISDL ETDVPDEDAD DEEEPLEIPR PLRALDQTPG
MouseRobo2b1(p) PPQPPAPPLG YVSGALISDL ETDVPDEDAD DEEEPLEIPR PLRALDQTPG
MouseRobo2b2(p) PPQPPAPPLG YVSGALISDL ETDVPDEDAD DEEEPLEIPR PLRALDQTPG
MouseRobo2b3(p) PPQPPAPPLG YVSGALISDL ETDVPDEDAD DEEEPLEIPR PLRALDQTPG
MouseRobo2b4(p) PPQPPAPPLG YVSGALISDL ETDVPDEDAD DEEEPLEIPR PLRALDQTPG


1301
1350
HUMANROBO2A(P) SSMDNLDSSV TGKAFTSSQR PRPTSPFSTD SNTSAALSQS QRPRPTKKHK
HUMANROBO2B(P) SSMDNLDSSV TGKAFTSSQR PRPTSPFSTD SNTSAALSQS QRPRPTKKHK
MouseRobo2a1(p) SSMDNLDSSV TGKAFSSSQR QRPTSPFSTD SNTSAAQNQS QRPRPTKKHK
MouseRobo2a2(p) SSMDNLDSSV TGKAFSSSQR QRPTSPFSTD SNTSAAQNQS QRPRPTKKHK
MouseRobo2a3(p) SSMDNLDSSV TGKAFSSSQR QRPTSPFSTD SNTSAAQNQS QRPRPTKKHK
MouseRobo2a4(p) SSMDNLDSSV TGKAFSSSQR QRPTSPFSTD SNTSAAQNQS QRPRPTKKHK
MouseRobo2b1(p) SSMDNLDSSV TGKAFSSSQR QRPTSPFSTD SNTSAAQNQS QRPRPTKKHK
MouseRobo2b2(p) SSMDNLDSSV TGKAFSSSQR QRPTSPFSTD SNTSAAQNQS QRPRPTKKHK
MouseRobo2b3(p) SSMDNLDSSV TGKAFSSSQR QRPTSPFSTD SNTSAAQNQS QRPRPTKKHK
MouseRobo2b4(p) SSMDNLDSSV TGKAFSSSQR QRPTSPFSTD SNTSAAQNQS QRPRPTKKHK


1351
1400
HUMANROBO2A(P) GGRMDQQPAL PHRREGMTDE EALVPYSKPS FPSPGGHSSS GTASSKGSTG
HUMANROBO2B(P) GGRMDQQPAL PHRREGMTDE EALVPYSKPS FPSPGGHSSS GTASSKGSTG
MouseRobo2a1(p) GGQMDPQPVL PHRREGMPDD LPPPPDPPPG QGLRQQIGLS QHSGNVENST
MouseRobo2a2(p) GGQMDPQPVL PHRREGMPDE ESLVPYSKPS FPSPGGHSSS GTSSSKGSTG
MouseRobo2a3(p) GGQMDPQPVL PHRREGMPDD LPPPPDPPPG QGLRQQIGLS QHSGNVENST
MouseRobo2a4(p) GGQMDPQPVL PHRREGMPDE ESLVPYSKPS FPSPGGHSSS GTSSSKGSTG
MouseRobo2b1(p) GGRMDPQPVL PHRREGMPDD LPPPPDPPPG QGLRQQIGLS QHSGNVENST
MouseRobo2b2(p) GGRMDPQPVL PHRREGMPDE ESLVPYSKPS FPSPGGHSSS GTSSSKGSTG
MouseRobo2b3(p) GGRMDPQPVL PHRREGMPDD LPPPPDPPPG QGLRQQIGLS QHSGNVENST
MouseRobo2b4(p) GGRMDPQPVL PHRREGMPDE ESLVPYSKPS FPSPGGHSSS GTSSSKGSTG




Blue: All amino acids of a column are identical.
Red: More than half of the amino acids of a column are identical or belong to one of the strong groups (amino acids with strong similarities).
Yellow: More than half of the amino acids of a column belong to one of the weak groups (amino acids with weak similarities) or marked are amino acids that could be grouped into a weak group with every amino acid of the same column belong to a strong group that marked red.

**Supplementary figure 3. Human *ROBO2a* sequence amplifed by primer pair No.2 + No 1.**

No.2

CTCCCAGACAGAGAGTGGGAGAAACCGGGGAGCAGCGGGAGCAGCAGGTCCGGGGGGAGCTGTTCCGCTGCGCTGCCCTCGTTATTCACAC
GGACGCTGCGGAGCTTCCCAGGGCTGCTTCCCTGTCCCCCTGGGTGGAGGCTGCCGTCTAAACCTGACTCCAGAGTTTAAGATGCAATGGC
CAGAAGACATGAACGTGTCACTAGAAGGATGTGGACATGGGCTCCGGGACTGTTGATGATGACTGTGGTGTTTTGGGGTCATCAGGGGAAT
GGACAAGGCCAAGGATCGCGTCTTCGCCAGGAGGACTTTCCCCCGCGGATTGTGGAGCATCCTTCCGATGTCATCGTCTCTAAGGGCGAGC
CCACGACTCTGAACTGCAAGGCGGAGGGCCGGCCAACGCCCACCATTGAGTGGTACAAAGATGGGGAGCGAGTGGAGACTGACAAGGACGA
TCCCCGGTCCCACAGGATGCTTCTGCCCAGCGGATCCTTATTCTTCTTGCGCATCGTGCACGGGCGCAGGAGTAAACCTGATGAAGGAAGC
TACGTTTGTGTTGCGAGGAACTATCTTGGTGAAGCAGTGAGTCGAAATGCGTCTCTGGAAGTGGCATTGTTACGAGATGACTTCCGACAAA
ACCCCACAGATGTTGTAGTGGCAGCTGGAGAGCCTGCAATCCTGGAGTGCCAGCCTCCCCGGGGACACCCAGAACCCACCATCTACTGGAA
AAAAGACAAAGTTCGAATTGATGACAAGGAAGAAAGAATAAGTATCCGTGGTGGAAAACTGATGATCTCCAATACCAGGAAAAGTGATGCA
GGGATGTATACTTGTGTTGGTACCAATATGGTGGGAGAAAGGGACAGTGACCCAGCAGAGCTGACTGTCTTTGAACGACCCACATTTCTCA
GGAGGCCAATTAACCAGGTGGTACTGGAGGAAGAAGCTGTAGAATTTCGTTGTCAAGTCCAAGGAGATCCTCAACCAACTGTGAGGTGGAA
AAAGGATGATGCAGACTTGCCAAGAGGAAGGTATGACATCAAAGACGATTACACACTAAGAATTAAAAAGACCATGAGTACAGATGAAGGC
ACCTATATGTGTATTGCTGAGAATCGGGTTGGAAAAATGGAAGCCTCTGCTACACTCACCGTCCGAGCTCCCCCACAGTTTGTGGTTCGGC
CAAGAGATCAGATTGTTGCTCAAGGTCGAACAGTGACATTTCCCTGTGAAACTAAAGGAAACCCACAGCCAGCTGTTTTTTGGCAGAAAGA
AGGCAGCCAGAACCTACTTTTCCCAAACCAACCCCAGCAGCCCAACAGTAGATGCTCAGTGTCACCAACTGGAGACCTCACAATCACCAAC
ATTCAACGTTCCGACGCGGGTTACTACATCTGCCAGGCTTTAACTGTGGCAGGAAGCATTTTAGCAAAAGCTCAACTGGAGGTTACTGATG
TTTTGACAGATAGACCTCCACCTATAATTCTACAAGGCCCAGCCAACCAAACGCTGGCAGTGGATGGTACAGCGTTACTGAAATGTAAAGC
CACTGGTGATCCTCTTCCTGTAATTAGCTGGTTAAAGGAGGGATTTACTTTTCCGGGTAGAGATCCAAGAGCAACAATTCAAGAGCAAGGC

No.1

ACACTGCAGATTAAGAATTTACGGATTTCTGATACTGGCACTTATACTTGTGTGGCTACA

Positions of primers are marked in gray.

**Supplementary figure 4. Human *ROBO2b* sequence amplifed by primer pair No.3 + No.1 and 5'RACE result with primer No.17.**

CTCCTCTCTTTTGGAAACCGGAGAGGTGGAGGGAGGGCAACACCGCTGCAAAGGAGAGGCCCGCCAAGTCTGCCCGCCTGCAAAGTGTTGC
TTTGACACATTCTTATTATGGAAGTTAAGTAAAAATATAGACATATTAAAAAATAACTCCGGACGTGGAGCTGCTACGGAGAAGGAAACCG
GGGGAAAGAAAACCAGTAGGCAGGCCAATGGTTTTTCGGCAGCGCGCTGGCAAGTTTGTGGAACACTTTCTAGGAATTAGGTCTTTTCCTC
CCCCTTCATCATCTTGACTTCTGAAGGAAGAACTTGGCTTTGGATTGCAGTGGAGCCTAAGGAGAGAGGGTTAGACACACTCGAATAATCC
CTCTGGCTGGGCTGAATTTGTGGGAATTTAGGAAGCCAGAGTGCTGGAAATACAGCAGCCTTTGAAGTACCCTCTGTTAATTTGGATGGAT
CTCAGTGTGCCCCGTTCGAGACCTCTCCACCAACCCCTTCTGATCTTGCGATTTGCTCTTCTTGACTTTAATTAGTATCTAGGAAAGTCTA
AACTTTGGACCTACCTCTTTTTTTGATACTCATTTTTGTACTTTTGCTCTCTGGGATTGGTTTCTTAAAGAATCTGGATCCTTTTTAATAT

No.3 No.17

GTCAAAATGAGTCTGCTGATGTTTACACAACTACTGCTCTGTGGATTTTTATATGTTCGGGTTGATGGATCGCGTCTTCGCCAGGAGGACT
TTCCCCCGCGGATTGTGGAGCATCCTTCCGATGTCATCGTCTCTAAGGGCGAGCCCACGACTCTGAACTGCAAGGCGGAGGGCCGGCCAAC
GCCCACCATTGAGTGGTACAAAGATGGGGAGCGAGTGGAGACTGACAAGGACGATCCCCGGTCCCACAGGATGCTTCTGCCCAGCGGATCC
TTATTCTTCTTGCGCATCGTGCACGGGCGCAGGAGTAAACCTGATGAAGGAAGCTACGTTTGTGTTGCGAGGAACTATCTTGGTGAAGCAG
TGAGTCGAAATGCGTCTCTGGAAGTGGCATTGTTACGAGATGACTTCCGACAAAACCCCACAGATGTTGTAGTGGCAGCTGGAGAGCCTGC
AATCCTGGAGTGCCAGCCTCCCCGGGGACACCCAGAACCCACCATCTACTGGAAAAAAGACAAAGTTCGAATTGATGACAAGGAAGAAAGA
ATAAGTATCCGTGGTGGAAAACTGATGATCTCCAATACCAGGAAAAGTGATGCAGGGATGTATACTTGTGTTGGTACCAATATGGTGGGAG
AAAGGGACAGTGACCCAGCAGAGCTGACTGTCTTTGAACGACCCACATTTCTCAGGAGGCCAATTAACCAGGTGGTACTGGAGGAAGAAGC
TGTAGAATTTCGTTGTCAAGTCCAAGGAGATCCTCAACCAACTGTGAGGTGGAAAAAGGATGATGCAGACTTGCCAAGAGGAAGGTATGAC
ATCAAAGACGATTACACACTAAGAATTAAAAAGACCATGAGTACAGATGAAGGCACCTATATGTGTATTGCTGAGAATCGGGTTGGAAAAA
TGGAAGCCTCTGCTACACTCACCGTCCGAGCTCCCCCACAGTTTGTGGTTCGGCCAAGAGATCAGATTGTTGCTCAAGGTCGAACAGTGAC
ATTTCCCTGTGAAACTAAAGGAAACCCACAGCCAGCTGTTTTTTGGCAGAAAGAAGGCAGCCAGAACCTACTTTTCCCAAACCAACCCCAG
CAGCCCAACAGTAGATGCTCAGTGTCACCAACTGGAGACCTCACAATCACCAACATTCAACGTTCCGACGCGGGTTACTACATCTGCCAGG
CTTTAACTGTGGCAGGAAGCATTTTAGCAAAAGCTCAACTGGAGGTTACTGATGTTTTGACAGATAGACCTCCACCTATAATTCTACAAGG
CCCAGCCAACCAAACGCTGGCAGTGGATGGTACAGCGTTACTGAAATGTAAAGCCACTGGTGATCCTCTTCCTGTAATTAGCTGGTTAAAG
GAGGGATTTACTTTTCCGGGTAGAGATCCAAGAGCAACAATTCAAGAGCAAGGCACACTGCAGATTAAGAATTTACGGATTTCTGATACTG

No.1

GCACTTATACTTGTGTGGCTACA

Positions of primers are marked in gray.

## Supplementary figure 5. Mouse *Robo2a* variant 1 sequence.

No.4
CTCTGTAGCTGACAGGCTGGGAGGAGAGGAAAGAAAACGAAGGAGGAGTGCTCTGGGCGGGGAGTGGAGAAACAAAAATCTGGGCGCATTT
AGAGCATCAGGTCCTGGGGGTGTTGTTCAGCTGCAGCGCACTCTTAATTCAGGCGGACCCTGCCTGAGTTGAGTTTCTCAGCAGATTGTCA
No.18
GTCCCTGTTGGAGCAGGCTGCCATCTAGACCTGACTCCTGAGTTTAAAATGCAATGGCCAGAAGACAAGAAAGTATCTGTGGAAGGCCGTG
GACGTGGACGCCAGGACTACTGATGCTGATCATCCTTGGCATTCATCAGGGGAGTGGGCAAGGTCAAGGATCTCGTCTTCGTCAAGAAGAC
TTTCCCCCCAGGATCGTAGAACACCCTTCTGATGTCATCGTCTCCAAGGGAGAGCCCACCACTCTGAACTGTAAAGCAGAGGGCCGACCCA
CCCCCACCATTGAATGGTACAAGGATGGTGAGAGGGTGGAGACAGACAAGGATGATCCCAGGTCCCACAGAATGCTTCTGCCCAGCGGATC
TTTATTCTTTTTGCGAATTGTTCATGGGCGCAGAAGTAAACCGGACGAAGGGAGTTATGTTTGTGTTGCAAGGAACTATCTTGGTGAAGCA
No.5
GTGAGTCGAAATGCATCTCTGGAAGTGGCATTATTGCGAGATGACTTCCGGCAAAACCCCACAGATGTGGTAGTCGCAGCTGGAGAGCCTG
CAATCTTGGAGTGCCAGCCACCACGGGGACACCCAGAACCAACCATCTACTGGAAAAAGGACAAAGTCCGAATTGATGaCAAGGAAGAGAG
AATAAGTATCCGTGGTGGGAAGCTGaTGATCTCTAATACTAGGAAAAGCGATGCTGGCATGTACACCTGTGTGGGAACCAATATGGTGGGA
GAAAGGGACAGCGACCCTGCAGAGCTCACTGTCTTTGAACGACCCACATTTCTCAGGAGGCCAATTAACCAGGTGGTGCTAGAGGAAGAAG
CTGTAGAGTTCCGTTGTCAGGTCCAAGGAGATCCCCAGCCAACGGTGAGGTGGAAAAAAGATGATGCAGACTTGCCGAGAGGAAGGTACGA
TATCAAAGATGACTACaCGCTGAGAATTAAAAAGGCCATGAGTACTGATGAAGGTACCTATGTGTGTATTGCTGAGAATCGGGTGGGAAAA
No.6
GTGGAAGCCTCTGCTACCCTCACTGTCCGAGTTCGCCCTGTTGCTCCTCCACAGTTTGTGGTTAGGCCAAGAGATCAGATCGTTGCTCAAG
GCCGAACAGTGACATTCCCCTGTGAAACTAAAGGAAACCCACAGCCAGCTGTTTTTTGGCAGAAAGAAGGCAGCCAGAACCTACTTTTCCC
GAATCAACCTCAGCAGCCCAACAGCCGATGTTCAGTGTCGCCCACGGGGGACCTCACCATCACCAACATCCAGCGTTCAGATGCGGGTTAC
TACATCTGCCAGGCCCTAACCGTGGCAGGAAGCATTTTAGCTAAAGCACAGTTGGAAGTTACTGACGTTTTGACAGATAGACCTCCACCCA
TAATCTTGCAAGGACCAATAAACCAAACACTTGCAGTAGACGGTACAGCATTGTTGAAGTGTAAAGCCACTGGTGAGCCTCTGCCTGTAAT
TAGCTGGCTAAAGGAGGGCTTTACTTTTCTGGGGAGAGATCCAAGAGCCACGATCCAAGACCAAGGAACACTGCAGATTAAGAATTTACGG
No.1 (rev) + No.7 (for)
ATATCTGATACTGGCACTTATACTTGTGTGGCTACAAGTTCCAGTGGAGAGACTTCCTGGAGTGCAGTGCTGGATGTAACAGAATCTGGAG
CAACAATCAGTAAAAATTATGATATGAATGACCTCCCGGGACCACCATCCAAACCTCAGGTCACTGATGTTTCTAAGAACAGTGTCACCTT
No.9
ATCCTGGCAGCCAGGTACACCTGGCGTTCTTCCTGCAAGCGCGTATATCATTGAGGCTTTCAGCCAATCGGTGAGCAATAGCTGGCAGACA
GTGGCAAACCATGTTAAGACAACTCTGTATACAGTAAGGGGGCTGAGGCCCAACACAATCTACTTGTTTATGGTCAGAGCGATCAACCCAC
AAGGTCTCAGTGATCCAAGTCCTATGTCGGATCCTGTACGCACACAAGATATCAGCCCCCCAGCACAAGGAGTGGACCACAGACAGGTACA
GAAGGAATTAGGTGATGTCGTTGTTCGTCTCCATAATCCAGTTGTCCTGACACCTACAACTGTTCAAGTCACATGGACGGTGGACCGACAA
CCCCAGTTTATTCAGGGCTACAGAGTGATGTACCGTCAGaCTTCGGGACTACAAGCCTCAACTGTGTGGCAGAATCTAGACGCcAAAGTCC
CGACTGAGAGGAGTGCTGTCCTTGTGAATTTGAAAAAGGGGGTGACTTATGAAATTAAAGTCCGGCCGTATTTTAACGAGTTCCAAGGAAT
GGACAGTGAATCGAAAACAGTCCGAACCACTGAGGAAGCCCCAAGTGCCCCTCCCCAGTCTGTCACTGTGCTGACAGTTGGAAGTCACAAC
AGCACAAGCATCAGTGTTTCCTGGGATCCTCCACCAGCCGACCACCAGAATGGAATTATTCAGGAATATAAGATCTGGTGTCTGGGAAACG
No.10
AAACGCGATTCCATATCAATAAAACGGTGGATGCAGCCATTCGCTCTGTAGTAATAGGTGGCTTGTTCCCTGGAATTCAGTACCGGGTAGA
AGTGGCAGCTAGCACAAGTGCAGGGGTTGGAGTAAAAAGTGAACCACAGCCGATAATAATTGGGGGACGTAACGAAGTTGTCATTACTGAA
AACAATAACAGCATCACTGAGCAAATCACGGATGTCGTGAAGCAACCGGCATTTATAGCTGGCATTGGTGGTGCCTGCTGGGTAATTCTGA
No.11 (for) + No.8 (rev)
TGGGTTTTAGCATCTGGTTGTACTGGAGAAGAAAGAAGAGAAAGGGACTCAGCAATTATGCTGTAACATTTCAAAGAGGAGATGGAGGACT
AATGAGCAATGGGAGCCGTCCAGGTCTTCTAAATGCTGGCGATCCCAATTACCCATGGCTTGCTGATTCTTGGCCAGCCACGAGTTTGCCA

GTGAACAATAGCAATAGTGGCCCAAATGAAATTGGAAATTTTGGGCGTGGAGATGTGCTGCCTCCGGTGCCAGGCCAAGGGGATAAAACAG
CGACCATGCTCTCGGATGGAGCCATTTATAGCAGCATTGACTTCACTACCAAAACCACTTACAACAGTTCCAGCCAAATAACACAGGCCAC
CCCATATGCCACTACACAAATCCTGCATTCAAACAGCATCCACGAACTGGCAGTTGATCTTCCTGATCCACAGTGGAAAAGCTCAGTTCAA
CAGAAGACAGACCTCATGGGATtTGGTTATTCGCTACCTGATCAGAACAAGGGGAACAACgGTGGGAAAGGTGGAAAAAAGAAGAAAACTA
AAAATTCTTCGAAAGCGCAGAAAAACAACGGATCCACTTGGGCTAATGTCCCTCTACCTCCTCCTCCCGTCCAGCCCCTTCCTGGTACAGA
GCTGGGCCACTATGCTGCGGAACAAGAAAATGGCTATGACAGTGACAGCTGGTGCCCACCATTACCGGTGCAAACATACTTGCATCAGGGT

No.13 (for) + No.12 (rev)

ATGGAAGATGAGCTGGAAGAAGACGAAGATCGGGTCCCAACACCTCCTGTTCGCGGCGTGGCCTCTTCACCAGCTATCTCTTTTGGACAGC
AGTCCACTGCCACTCTTACTCCATCCCCACGGGAAGAGATGCAACCCATGCTGCAAGCTCACTTGGATGAGTTGACAAGGGCCTATCAGTT
TGATATAGCAAAACAAACATGGCACATTCAAAGCAATACCCCACCTCCACAACCCCCAGCTCCGCCATTAGGTTACGTGTCCGGAGCCTTG
ATTTCTGATTTGGAGACAGATGTTCCAGATGAGGATGCTGATGATGAAGAGGAACCATTAGAAATTCCCAGGCCCCTCAGAGCACTAGACC
AGACACCTGGATCCAGTATGGACAATCTAGACAGCTCTGTCACAGGAAAAGCCTTTAGCTCCTCTCAAAGGCAGCGGCCCACCAGCCCATT
TTCTACGGACAGTAACACCAGTGCTGCCCAGAATCAAAGCCAGAGGCCTCGGCCCACGAAAAAACATAAGGGAGGACAGATGGACCCACAG
CCAGTGTTACCTCATCGGAGGGAAGGGATGCCAGATGATCTTCCCCCACCACCAGATCCCCCACCAGGTCAGGGTTTAAGGCAGCAGATAG
GCCTGAGCCAACACTCTGGTAACGTGGAAAACTCAACAGAGAGAAAAGGAAGCTCTCTAGAGAGACAACAAGCAGCCAACTTAGAAGACAC
AAAGAGCTCATTGGATTGTCCAGCTAAAACTGTTCTAGAGTGGCAGCGTCAAACCCAGGACTGGATAAACTCTACAGAACGCCAAGAAGAA

No.15

ACTCGGAAAGCCCCACACAAACAAGGGGTTGGATCAGAGGAGTCCCTGGTGCCCTACAGCAAGCCCAGCTTCCCGTCTCCTGGCGGGCACA
GCTCATCAGGAACATCTTCCTCTAAAGGGTCCACTGGCCCTCGGAAGGCTGACGTGCTGAGAGGCAGCCACCAGAGGAATGCCAATGACCT
TCTCGACATTGGATATGTGGGCTCAAATAGTCAAGGACAGTTTACAGAGTAACTGAGAGGGGACATACAAAGATGCTCTGAGGACCATCAG
GTCCAAACTCATGGAAGTGATAACATTTAACAGTGCAATGAACAATTTCTTTATTTATGTACTATTAAAAAAAAACTGTAAATGCAATGTA
AAGACACACAGCCACACATATTCCACAGTTATTTTCTTGTGTTCTTCTCTTACTTACACCACCCACCTTAACTCTTTCTTGTCAGGAGTAT
ATAAGAAAAAGAAAGAAAGCAAAACTCACCCTCCAGGAAGAAAAGGATTTTTCCTCTGTATATAATTTCTTTTGTGCATTGCTATGCAAGC
TCACTCTTTTTAACTCTGTTCATATTATTGTCTGTTCTGATTGGTCTGTTGTACTATATGTGAATTAATGGGCTGTGGTGCCATATATTAA
CTTTTAATTGTGTAACTTTTTATGTTTAAATTTTGCACTGCGATTTTATTTGGTGATAAGCACAAATCTCTACTCCTCATGACATGAAGAA

No.16

AAGATTGAATGTGAAGGGAGTTTTCTGTACTGTAAGTTAGGTTGGATAATGCTGGTGTAACCAATCCAGTTAGATGGTTTTCAGTTGGGGG
TGTAGAAATAGGAAGATCGAAGGAATGATGGTGTTGGCAAAGTCTTCTTGAAACAACAGATATTGAGACAATTTTAAGAAGCAGAAAGATG
GATACTATTGACTAAAGCAGGGGTCAAAAGAAGGGGGTTTAAGTCTAGACAGAGTATGTAATAAAGTATGGTGGTAGCAAAGATGTACTAA
CTTGCTTTAAAAATATTATTAAAGTTTTATTTAGAATGAACTTTACCTGCCATTGTAATTAACCCATCTTAGAATTACAATGAGCAAAATA
AAAATCAAGGTGTTTCAAGAGATCTGTATTTATATTACAGCTTTTTAAAAACAGCAGTTTTGAATAATCATAATTAATCTCTCCAACTCAT

No.14

TAAGTCAGACTATAACATGGAGTTCCCCAAGGAAACAAACAAAAAAAATGAACTCCAGAAACTCTAGCAAATAGTTAAAGAAGCAACTTAT
TCTTCGAGCATACTCGGCTACTCTTGAATATGAGCTCAATTGCAATATCTTGTTCATCTTTCTAATACATGTACAGTACATATTAGAGGAT
GTAGGCACATCACTTAAGTTCATTTCTTTTTAAAAAAAAAAAGCAAACTAATGCATTCTGCATACAATTTTGTGTTTCTTTTGGTTAAAAT
GTGAAATCGATGCCACCAGATGTATAGCCACATTATAAGTGCTTAAAGGCAGTGCTCACAGTATGTTCAGTTTCCAGTTAGTAGATTTATT
GGACTCAATATTTTCTGGTACATTCTCAGAAATTGGCTACCAACTCAAATCTGTTTGACCCATTTTGAATGAGTAAATCGGAAAGAACAGT
CAAAAGGGAGAAAAAAATGCATGTTTATACACTTTTTAAATAAAACCAAATCTTGTAAGTGGACATTAATAACAGTGTCCTGCCTCATTTG
TTTTCACGACTCTGAGAACAAGGAGTTGCTCAACATCCATTATGTATGCTCAAGATAAATGTGACTTTGAAATGTATGTTAGCTACTGTAC
ATGTATGTAAGTCAAGTAGATGATAGCCTTCCTGAAACTACCACTAATAATATTTTCCCATGTGCTACTTACACAAGCTTAAATTCTAGTT
CAGCAAATGTTTTTCATTATTTTTTCTTACGGTTTATTCAATAGAGGGTACCTAGAAAAATATTCAGTGTGATGGTCAACATATTGAGTGT
TCAAGTAGGGGATATTTACTTTGGCAATACATACTGTGCATTATTGGTTCAGATGCCAATTCATGAGACATTATCATAGGAAATGGATAAT
ATAAGAACATGGCTCTAAGGGTCCTTCTCCACAGGCAAGCATAGAATTTGAGGTTCTTCATGGGTCCCAAAACTTGCTTAGTGTTATATAC
TTAACTTCCTTCACTGAAGTATCAAGCAATTCTAACACACCGTTTAACAACTTTCTACCATCAGTGCCTGCATTTTATTGAAGTTTGATGT

CTCTGGTACAAGGAAATAGAGAAAGCTGAACAGTTTCTGTTGGGAGAAAAAGTTTCTGTAATTATTGTCAATTCTAAGTTGTGATGAATAT
AACTTTTCATCTATACAGTTCCTTACTCATGGTCTCTATTTAAAAGCATAGGATCTTTTAAGATATACTTTTAGCTGAGTCTTCACTGTAT
GTTCATGAATAATATAATTTAGTGAAAAACTTTAGTACATCATATGTATGGTAACTGACATTTGCAGATTATTATCAAACCAAATCATAAG
AAATAATATCTTCAGATCACATGATGGCTTACCGAACAATTTCAGGTTATAGCATGATGTCTTTCTGATTGTCTTCAGTATATGATCATAA
GCAATTATATTGTGTCCATACAGTTTACATACTGTGATTCCCGCAAAATCTTGCATCTGATATCTACCATGAAGATGAAATAATATGCTAT
GAATTTCAGGCATAGACATACTATCAGCAATAGAAAGCCTTCATCATCTAGATATCTGGATAAATCCATTCATCGTCAACCATCAGATTGT
TAACTGGGGAACACGTACCTCCCTCACAAATGCTTTCTTTTCGTCTTGTAGTTGATTGAATGTGATAGGAAATCTGCTGAGTTGACTCTTC
TGTGGGCCTGAGCAGAAACATGCATAAGCCATCAGAAGGAAGTCTGGATGGGAAGGAAACCTTGTCTTTCACTCGTAGGCAATACTGCTGT
GTAGTCACGGCAGACCCAAGCACAGCTTCTTTCAGAACATGCAAAGTCTCCTTATGAAATATACACATCAGTGCCCAAACTCAGGTGTACC
AGTGTAAAGCGTAGGGTTTGCCTACCCTTTTTTCATATCATTTTACATAAGCAAGTTAGGAGACTACAGGGCTATGAAAATAAATGTTTCC
TTAACAGGATTTGGAGCAAGTTCACGGGAGTACATAGGACAGTCGAGCCCAGTGTAAACACTACCAAAGGAAAGGACAGCCTTGCAGAAGC
GTGAGCTCCGACTAACAGAGCGGCTTAAGTATGCTCTGTGATACAGCTTAATGCACTCCCTATTCAAATCCCACTGTCAATGGGATTTAAG
TTGCCAGTTTATATGTGATTTCATGGAACACAACCAAGTTCTACTTAATACCTTTATGCAATTTTTTTTTCACTGATGAAAAATCAATATC
CACATCAATGGGTGCCCAGCTGCGGATCACCTCTCCTAAAAAGCGATCTTGTCAGTATTGAAGAAATGTAGAACCATAATTTTCACTACTA
ATATCATTTTGCTCAACTACTTCTGTATATTGTGGAAAAAGGGGTGGGGCCGGGGGAATGTGTGTTTCCACATTTTCTGAGAGGTGAAAAT
ATTTTTGTGTTGCTTTGTCCTAATGGGTACTTCAAATGAAACACCATACAATTTTAAATGACATGCAATTAACGAATATTTTTTACTCTCA
TAACATTAAAAGCATTTGTTGTG

Positions of primers are marked in gray.

## Supplementary figure 6. Mouse *Robo2a* variant 1 sequence.

No.4

CTCTGTAGCTGACAGGCTGGGAGGAGAGGAAAGAAAACGAAGGAGGAGTGCTCTGGGCGGGGAGTGGAGAAACAAAAATCTGGGCGCATTT
AGAGCATCAGGTCCTGGGGGTGTTGTTCAGCTGCAGCGCACTCTTAATTCAGGCGGACCCTGCCTGAGTTGAGTTTCTCAGCAGATTGTCA
GTCCCTGTTGGAGCAGGCTGCCATCTAGACCTGACTCCTGAGTTTAAAATGCAATGGCCAGAAGACAAGAAAGTATCTGTGGAAGGCCGTG
GACGTGGACGCCAGGACTACTGATGCTGATCATCCTTGGCATTCATCAGGGGAGTGGGCAAGGTCAAGGATCTCGTCTTCGTCAAGAAGAC
TTTCCCCCCAGGATCGTAGAACACCCTTCTGATGTCATCGTCTCCAAGGGAGAGCCCACCACTCTGAACTGTAAAGCAGAGGGCCGACCCA
CCCCCACCATTGAATGGTACAAGGATGGTGAGAGGGTGGAGACAGACAAGGATGATCCCAGGTCCCACAGAATGCTTCTGCCCAGCGGATC
TTTATTCTTTTTGCGAATTGTTCATGGGCGCAGAAGTAAACCGGACGAAGGGAGTTATGTTTGTGTTGCAAGGAACTATCTTGGTGAAGCA

No.5

GTGAGTCGAAATGCATCTCTGGAAGTGGCATTATTGCGAGATGACTTCCGGCAAAACCCCACAGATGTGGTAGTCGCAGCTGGAGAGCCTG
CAATCTTGGAGTGCCAGCCACCACGGGGACACCCAGAACCAACCATCTACTGGAAAAAGGACAAAGTCCGAATTGATGaCAAGGAAGAGAG
AATAAGTATCCGTGGTGGGAAGCTGaTGATCTCTAATACTAGGAAAAGCGATGCTGGCATGTACACCTGTGTGGGAACCAATATGGTGGGA
GAAAGGGACAGCGACCCTGCAGAGCTCACTGTCTTTGAACGACCCACATTTCTCAGGAGGCCAATTAACCAGGTGGTGCTAGAGGAAGAAG
CTGTAGAGTTCCGTTGTCAGGTCCAAGGAGATCCCCAGCCAACGGTGAGGTGGAAAAAAGATGATGCAGACTTGCCGAGAGGAAGGTACGA
TATCAAAGATGACTACaCGCTGAGAATTAAAAAGGCCATGAGTACTGATGAAGGTACCTATGTGTGTATTGCTGAGAATCGGGTGGGAAAA

No.6

GTGGAAGCCTCTGCTACCCTCACTGTCCGAGTTCGCCCTGTTGCTCCTCCACAGTTTGTGGTTAGGCCAAGAGATCAGATCGTTGCTCAAG
GCCGAACAGTGACATTCCCCTGTGAAACTAAAGGAAACCCACAGCCAGCTGTTTTTTGGCAGAAAGAAGGCAGCCAGAACCTACTTTTCCC
GAATCAACCTCAGCAGCCCAACAGCCGATGTTCAGTGTCGCCCACGGGGGACCTCACCATCACCAACATCCAGCGTTCAGATGCGGGTTAC
TACATCTGCCAGGCCCTAACCGTGGCAGGAAGCATTTTAGCTAAAGCACAGTTGGAAGTTACTGACGTTTTGACAGATAGACCTCCACCCA
TAATCTTGCAAGGACCAATAAACCAAACACTTGCAGTAGACGGTACAGCATTGTTGAAGTGTAAAGCCACTGGTGAGCCTCTGCCTGTAAT
TAGCTGGCTAAAGGAGGGCTTTACTTTTCTGGGGAGAGATCCAAGAGCCACGATCCAAGACCAAGGAACACTGCAGATTAAGAATTTACGG

No.1(rev)+No.7(for)

ATATCTGATACTGGCACTTATACTTGTGTGGCTACAAGTTCCAGTGGAGAGACTTCCTGGAGTGCAGTGCTGGATGTAACAGAATCTGGAG
CAACAATCAGTAAAAATTATGATATGAATGACCTCCCGGGACCACCATCCAAACCTCAGGTCACTGATGTTTCTAAGAACAGTGTCACCTT

No.9

ATCCTGGCAGCCAGGTACACCTGGCGTTCTTCCTGCAAGCGCGTATATCATTGAGGCTTTCAGCCAATCGGTGAGCAATAGCTGGCAGACA
GTGGCAAACCATGTTAAGACAACTCTGTATACAGTAAGGGGGCTGAGGCCCAACACAATCTACTTGTTTATGGTCAGAGCGATCAACCCAC
AAGGTCTCAGTGATCCAAGTCCTATGTCGGATCCTGTACGCACACAAGATATCAGCCCCCCAGCACAAGGAGTGGACCACAGACAGGTACA
GAAGGAATTAGGTGATGTCGTTGTTCGTCTCCATAATCCAGTTGTCCTGACACCTACAACTGTTCAAGTCACATGGACGGTGGACCGACAA
CCCCAGTTTATTCAGGGCTACAGAGTGATGTACCGTCAGaCTTCGGGACTACAAGCCTCAACTGTGTGGCAGAATCTAGACGCcAAAGTCC
CGACTGAGAGGAGTGCTGTCCTTGTGAATTTGAAAAAGGGGGTGACTTATGAAATTAAAGTCCGGCCGTATTTTAACGAGTTCCAAGGAAT
GGACAGTGAATCGAAAACAGTCCGAACCACTGAGGAAGCCCCAAGTGCCCCTCCCCAGTCTGTCACTGTGCTGACAGTTGGAAGTCACAAC
AGCACAAGCATCAGTGTTTCCTGGGATCCTCCACCAGCCGACCACCAGAATGGAATTATTCAGGAATATAAGATCTGGTGTCTGGGAAACG

No.10

AAACGCGATTCCATTCAATAAAACGGTGGATGCAGCCATTCGCTCTGTAGTAATAGGTGGCTTGTTCCCTGGAATTCAGTACCGGGTAGAA
GTGGCAGCTAGCACAAGTGCAGGGGTTGGAGTAAAAAGTGAACCACAGCCGATAATAATTGGGGGACGTAACGAAGTTGTCATTACTGAAA
ACAATAACAGCATCACTGAGCAAATCACGGATGTCGTGAAGCAACCGGCATTTATAGCTGGCATTGGTGGTGCCTGCTGGGTAATTCTGAT

No.11(for)+ No.8(rev)

GGGTTTTAGCATCTGGTTGTACTGGAGAAGAAAGAAGAGAAAGGGACTCAGCAATTATGCTGTAACATTTCAAAGAGGAGATGGAGGACTA
ATGAGCAATGGGAGCCGTCCAGGTCTTCTAAATGCTGGCGATCCCAATTACCCATGGCTTGCTGATTCTTGGCCAGCCACGAGTTTGCCAG
TGAACAATAGCAATAGTGGCCCAAATGAAATTGGAAATTTTGGGCGTGGAGATGTGCTGCCTCCGGTGCCAGGCCAAGGGGATAAAACAGC

GACCATGCTCTCGGATGGAGCCATTTATAGCAGCATTGACTTCACTACCAAAACCACTTACAACAGTTCCAGCCAAATAACACAGGCCACC
CCATATGCCACTACACAAATCCTGCATTCAAACAGCATCCACGAACTGGCAGTTGATCTTCCTGATCCACAGTGGAAAAGCTCAGTTCAAC
AGAAGACAGACCTCATGGGATtTGGTTATTCGCTACCTGATCAGAACAAGGGGAACAACgGTGGGAAAGGTGGAAAAAAGAAGAAAACTAA
AAATTCTTCGAAAGCGCAGAAAAACAACGGATCCACTTGGGCTAATGTCCCTCTACCTCCTCCTCCCGTCCAGCCCCTTCCTGGTACAGAG
CTGGGCCACTATGCTGCGGAACAAGAAAATGGCTATGACAGTGACAGCTGGTGCCCACCATTACCGGTGCAAACATACTTGCATCAGGGTA

No.13(for)+ No.12(rev)

TGGAAGATGAGCTGGAAGAAGACGAAGATCGGGTCCCAACACCTCCTGTTCGCGGCGTGGCCTCTTCACCAGCTATCTCTTTTGGACAGCA
GTCCACTGCCACTCTTACTCCATCCCCACGGGAAGAGATGCAACCCATGCTGCAAGCTCACTTGGATGAGTTGACAAGGGCCTATCAGTTT
GATATAGCAAAACAAACATGGCACATTCAAAGCAATACCCCACCTCCACAACCCCCAGCTCCGCCATTAGGTTACGTGTCCGGAGCCTTGA
TTTCTGATTTGGAGACAGATGTTCCAGATGAGGATGCTGATGATGAAGAGGAACCATTAGAAATTCCCAGGCCCCTCAGAGCACTAGACCA
GACACCTGGATCCAGTATGGACAATCTAGACAGCTCTGTCACAGGAAAAGCCTTTAGCTCCTCTCAAAGGCAGCGGCCCACCAGCCCATTT
TCTACGGACAGTAACACCAGTGCTGCCCAGAATCAAAGCCAGAGGCCTCGGCCCACGAAAAAACATAAGGGAGGACAGATGGACCCACAGC

No.15

CAGTGTTACCTCATCGGAGGGAAGGGATGCCAGATGAGGAGTCCCTGGTGCCCTACAGCAAGCCCAGCTTCCCGTCTCCTGGCGGGCACAG
CTCATCAGGAACATCTTCCTCTAAAGGGTCCACTGGCCCTCGGAAGGCTGACGTGCTGAGAGGCAGCCACCAGAGGAATGCCAATGACCTT
CTCGACATTGGATATGTGGGCTCAAATAGTCAAGGACAGTTTACAGAGTAACTGAGAGGGGACATACAAAGATGCTCTGAGGACCATCAGG
TCCAAACTCATGGAAGTGATAACATTTAACAGTGCAATGAACAATTTCTTTATTTATGTACTATTAAAAAAAAACTGTAAATGCAATGTAA
AGACACACAGCCACACATATTCCACAGTTATTTTCTTGTGTTCTTCTCTTACTTACACCACCCACCTTAACTCTTTCTTGTCAGGAGTATA
TAAGAAAAAGAAAGAAAGCAAAACTCACCCTCCAGGAAGAAAAGGATTTTTCCTCTGTATATAATTTCTTTTGTGCATTGCTATGCAAGCT
CACTCTTTTTAACTCTGTTCATATTATTGTCTGTTCTGATTGGTCTGTTGTACTATATGTGAATTAATGGGCTGTGGTGCCATATATTAAC
TTTTAATTGTGTAACTTTTTATGTTTAAATTTTGCACTGCGATTTTATTTGGTGATAAGCACAAATCTCTACTCCTCATGACATGAAGAAA

No.16

AGATTGAATGTGAAGGGAGTTTTCTGTACTGTAAGTTAGGTTGGATAATGCTGGTGTAACCAATCCAGTTAGATGGTTTTCAGTTGGGGGT
GTAGAAATAGGAAGATCGAAGGAATGATGGTGTTGGCAAAGTCTTCTTGAAACAACAGATATTGAGACAATTTTAAGAAGCAGAAAGATGG
ATACTATTGACTAAAGCAGGGGTCAAAAGAAGGGGGTTTAAGTCTAGACAGAGTATGTAATAAAGTATGGTGGTAGCAAAGATGTACTAAC
TTGCTTTAAAAATATTATTAAAGTTTTATTTAGAATGAACTTTACCTGCCATTGTAATTAACCCATCTTAGAATTACAATGAGCAAAATAA
AAATCAAGGTGTTTCAAGAGATCTGTATTTATATTACAGCTTTTTAAAAACAGCAGTTTTGAATAATCATAATTAATCTCTCCAACTCATT

No.14

AAGTCAGACTATAACATGGAGTTCCCCAAGGAAACAAACAAAAAAAAATGAACTCCAGAAACTCTAGCAAATAGTTAAAGAAGCAACTTATT
CTTCGAGCATACTCGGCTACTCTTGAATATGAGCTCAATTGCAATATCTTGTTCATCTTTCTAATACATGTACAGTACATATTAGAGGATG
TAGGCACATCACTTAAGTTCATTTCTTTTTAAAAAAAAAAAGCAAACTAATGCATTCTGCATACAATTTTGTGTTTCTTTTGGTTAAAATG
TGAAATCGATGCCACCAGATGTATAGCCACATTATAAGTGCTTAAAGGCAGTGCTCACAGTATGTTCAGTTTCCAGTTAGTAGATTTATTG
GACTCAATATTTTCTGGTACATTCTCAGAAATTGGCTACCAACTCAAATCTGTTTGACCCATTTTGAATGAGTAAATCGGAAAGAACAGTC
AAAAGGGAGAAAAAAATGCATGTTTATACACTTTTTAAATAAAACCAAATCTTGTAAGTGGACATTAATAACAGTGTCCTGCCTCATTTGT
TTTCACGACTCTGAGAACAAGGAGTTGCTCAACATCCATTATGTATGCTCAAGATAAATGTGACTTTGAAATGTATGTTAGCTACTGTACA
TGTATGTAAGTCAAGTAGATGATAGCCTTCCTGAAACTACCACTAATAATATTTTCCCATGTGCTACTTACACAAGCTTAAATTCTAGTTC
AGCAAATGTTTTTCATTATTTTTTCTTACGGTTTATTCAATAGAGGGTACCTAGAAAAATATTCAGTGTGATGGTCAACATATTGAGTGTT
CAAGTAGGGGATATTTACTTTGGCAATACATACTGTGCATTATTGGTTCAGATGCCAATTCATGAGACATTATCATAGGAAATGGATAATA
TAAGAACATGGCTCTAAGGGTCCTTCTCCACAGGCAAGCATAGAATTTGAGGTTCTTCATGGGTCCCAAAACTTGCTTAGTGTTATATACT
TAACTTCCTTCACTGAAGTATCAAGCAATTCTAACACACCGTTTAACAACTTTCTACCATCAGTGCCTGCATTTTATTGAAGTTTGATGTC
TCTGGTACAAGGAAATAGAGAAAGCTGAACAGTTTCTGTTGGGAGAAAAAGTTTCTGTAATTATTGTCAATTCTAAGTTGTGATGAATATA
ACTTTTCATCTATACAGTTCCTTACTCATGGTCTCTATTTAAAAGCATAGGATCTTTTAAGATATACTTTTAGCTGAGTCTTCACTGTATG
TTCATGAATAATATAATTTAGTGAAAAACTTTAGTACATCATATGTATGGTAACTGACATTTGCAGATTATTATCAAACCAAATCATAAGA
AATAATATCTTCAGATCACATGATGGCTTACCGAACAATTTCAGGTTATAGCATGATGTCTTTCTGATTGTCTTCAGTATATGATCATAAG

CAATTATATTGTGTCCATACAGTTTACATACTGTGATTCCCGCAAAATCTTGCATCTGATATCTACCATGAAGATGAAATAATATGCTATG
AATTTCAGGCATAGACATACTATCAGCAATAGAAAGCCTTCATCATCTAGATATCTGGATAAATCCATTCATCGTCAACCATCAGATTGTT
AACTGGGGAACACGTACCTCCCTCACAAATGCTTTCTTTTCGTCTTGTAGTTGATTGAATGTGATAGGAAATCTGCTGAGTTGACTCTTCT
GTGGGCCTGAGCAGAAACATGCATAAGCCATCAGAAGGAAGTCTGGATGGGAAGGAAACCTTGTCTTTCACTCGTAGGCAATACTGCTGTG
TAGTCACGGCAGACCCAAGCACAGCTTCTTTCAGAACATGCAAAGTCTCCTTATGAAATATACACATCAGTGCCCAAACTCAGGTGTACCA
GTGTAAAGCGTAGGGTTTGCCTACCCTTTTTTCATATCATTTTACATAAGCAAGTTAGGAGACTACAGGGCTATGAAAATAAATGTTTCCT
TAACAGGATTTGGAGCAAGTTCACGGGAGTACATAGGACAGTCGAGCCCAGTGTAAACACTACCAAAGGAAAGGACAGCCTTGCAGAAGCG
TGAGCTCCGACTAACAGAGCGGCTTAAGTATGCTCTGTGATACAGCTTAATGCACTCCCTATTCAAATCCCACTGTCAATGGGATTTAAGT
TGCCAGTTTATATGTGATTTCATGGAACACAACCAAGTTCTACTTAATACCTTTATGCAATTTTTTTTTCACTGATGAAAAATCAATATCC
ACATCAATGGGTGCCCAGCTGCGGATCACCTCTCCTAAAAAGCGATCTTGTCAGTATTGAAGAAATGTAGAACCATAATTTTCACTACTAA
TATCATTTTGCTCAACTACTTCTGTATATTGTGGAAAAAGGGGTGGGGCCGGGGGAATGTGTGTTTCCACATTTTCTGAGAGGTGAAAATA
TTTTTGTGTTGCTTTGTCCTAATGGGTACTTCAAATGAAACACCATACAATTTTAAATGACATGCAATTAACGAATATTTTTTACTCTCAT
AACATTAAAAGCATTTGTTGTG

Positions of primers are marked in gray.

## Supplementary figure 7. Mouse *Robo2a* variant 3.

No.4

CTCTGTAGCTGACAGGCTGGGAGGAGAGGAAAGAAAACGAAGGAGGAGTGCTCTGGGCGGGGAGTGGAGAAACAAAAATCTGGGCGCATTT
AGAGCATCAGGTCCTGGGGGTGTTGTTCAGCTGCAGCGCACTCTTAATTCAGGCGGACCCTGCCTGAGTTGAGTTTCTCAGCAGATTGTCA
GTCCCTGTTGGAGCAGGCTGCCATCTAGACCTGACTCCTGAGTTTAAAATGCAATGGCCAGAAGACAAGAAAGTATCTGTGGAAGGCCGTG
GACGTGGACGCCAGGACTACTGATGCTGATCATCCTTGGCATTCATCAGGGGAGTGGGCAAGGTCAAGGATCTCGTCTTCGTCAAGAAGAC
TTTCCCCCCAGGATCGTAGAACACCCTTCTGATGTCATCGTCTCCAAGGGAGAGCCCACCACTCTGAACTGTAAAGCAGAGGGCCGACCCA
CCCCCACCATTGAATGGTACAAGGATGGTGAGAGGGTGGAGACAGACAAGGATGATCCCAGGTCCCACAGAATGCTTCTGCCCAGCGGATC
TTTATTCTTTTTGCGAATTGTTCATGGGCGCAGAAGTAAACCGGACGAAGGGAGTTATGTTTGTGTTGCAAGGAACTATCTTGGTGAAGCA

No.5

GTGAGTCGAAATGCATCTCTGGAAGTGGCATTATTGCGAGATGACTTCCGGCAAAACCCCACAGATGTGGTAGTCGCAGCTGGAGAGCCTG
CAATCTTGGAGTGCCAGCCACCACGGGGACACCCAGAACCAACCATCTACTGGAAAAAGGACAAAGTCCGAATTGATGACAAGGAAGAGAG
AATAAGTATCCGTGGTGGGAAGCTGATGATCTCTAATACTAGGAAAAGCGATGCTGGCATGTACACCTGTGTGGGAACCAATATGGTGGGA
GAAAGGGACAGCGACCCTGCAGAGCTCACTGTCTTTGAACGACCCACATTTCTCAGGAGGCCAATTAACCAGGTGGTGCTAGAGGAAGAAG
CTGTAGAGTTCCGTTGTCAGGTCCAAGGAGATCCCCAGCCAACGGTGAGGTGGAAAAAAGATGATGCAGACTTGCCGAGAGGAAGGTACGA
TATCAAAGATGACTACACGCTGAGAATTAAAAAGGCCATGAGTACTGATGAAGGTACCTATGTGTGTATTGCTGAGAATCGGGTGGGAAAA

No.6

GTGGAAGCCTCTGCTACCCTCACTGTCCGAGTTCGCCCTGTTGCTCCTCCACAGTTTGTGGTTAGGCCAAGAGATCAGATCGTTGCTCAAG
GCCGAACAGTGACATTCCCCTGTGAAACTAAAGGAAACCCACAGCCAGCTGTTTTTTGGCAGAAAGAAGGCAGCCAGAACCTACTTTTCCC
GAATCAACCTCAGCAGCCCAACAGCCGATGTTCAGTGTCGCCCACGGGGGACCTCACCATCACCAACATCCAGCGTTCAGATGCGGGTTAC
TACATCTGCCAGGCCCTAACCGTGGCAGGAAGCATTTTAGCTAAAGCACAGTTGGAAGTTACTGACGTTTTGACAGATAGACCTCCACCCA
TAATCTTGCAAGGACCAATAAACCAAACACTTGCAGTAGACGGTACAGCATTGTTGAAGTGTAAAGCCACTGGTGAGCCTCTGCCTGTAAT
TAGCTGGCTAAAGGAGGGCTTTACTTTTCTGGGGAGAGATCCAAGAGCCACGATCCAAGACCAAGGAACACTGCAGATTAAGAATTTACGG

No.1(rev)+ No.7(for)

ATATCTGATACTGGCACTTATACTTGTGTGGCTACAAGTTCCAGTGGAGAGACTTCCTGGAGTGCAGTGCTGGATGTAACAGAATCTGGAG
CAACAATCAGTAAAAATTATGATATGAATGACCTCCCGGGACCACCATCCAAACCTCAGGTCACTGATGTTTCTAAGAACAGTGTCACCTT

No.9

ATCCTGGCAGCCAGGTACACCTGGCGTTCTTCCTGCAAGCGCGTATATCATTGAGGCTTTCAGCCAATCGGTGAGCAATAGCTGGCAGACA
GTGGCAAACCATGTTAAGACAACTCTGTATACAGTAAGGGGGCTGAGGCCCAACACAATCTACTTGTTTATGGTCAGAGCGATCAACCCAC
AAGGTCTCAGTGATCCAAGTCCTATGTCGGATCCTGTACGCACACAAGATATCAGCCCCCCAGCACAAGGAGTGGACCACAGACAGGTACA
GAAGGAATTAGGTGATGTCGTTGTTCGTCTCCATAATCCAGTTGTCCTGACACCTACAACTGTTCAAGTCACATGGACGGTGGACCGACAA
CCCCAGTTTATTCAGGGCTACAGAGTGATGTACCGTCAGACTTCGGGACTACAAGCCTCAACTGTGTGGCAGAATCTAGACGCCAAAGTCC
CGACTGAGAGGAGTGCTGTCCTTGTGAATTTGAAAAAGGGGGTGACTTATGAAATTAAAGTCCGGCCGTATTTTAACGAGTTCCAAGGAAT
GGACAGTGAATCGAAAACAGTCCGAACCACTGAGGAAGCCCCAAGTGCCCCTCCCCAGTCTGTCACTGTGCTGACAGTTGGAAGTCACAAC
AGCACAAGCATCAGTGTTTCCTGGGATCCTCCACCAGCCGACCACCAGAATGGAATTATTCAGGAATATAAGATCTGGTGTCTGGGAAACG

No.10

AAACGCGATTCCATATCAATAAAACGGTGGATGCAGCCATTCGCTCTGTAGTAATAGGTGGCTTGTTCCCTGGAATTCAGTACCGGGTAGA
AGTGGCAGCTAGCACAAGTGCAGGGGTTGGAGTAAAAAGTGAACCACAGCCGATAATAATTGGGGGACGTAACGAAGTTGTCATTACTGAA
AACAATAACAGCATCACTGAGCAAATCACGGATGTCGTGAAGCAACCGGCATTTATAGCTGGCATTGGTGGTGCCTGCTGGGTAATTCTGA

No.11(for)+ No.8(rev)

TGGGTTTTAGCATCTGGTTGTACTGGAGAAGAAAGAAGAGAAAGGGACTCAGCAATTATGCTGTAACATTTCAAAGAGGAGATGGAGGACT
AATGAGCAATGGGAGCCGTCCAGGTCTTCTAAATGCTGGCGATCCCAATTACCCATGGCTTGCTGATTCTTGGCCAGCCACGAGTTTGCCA
GTGAACAATAGCAATAGTGGCCCAAATGAAATTGGAAATTTTGGGCGTGGAGATGTGCTGCCTCCGGTGCCAGGCCAAGGGGATAAAACAG

CGACCATGCTCTCGGATGGAGCCATTTATAGCAGCATTGACTTCACTACCAAAACCACTTACAACAGTTCCAGCCAAATAACACAGGCCAC
CCCATATGCCACTACACAAATCCTGCATTCAAACAGCATCCACGAACTGGCAGTTGATCTTCCTGATCCACAGTGGAAAAGCTCAGTTCAA
CAGAAGACAGACCTCATGGGATTTGGTTATTCGCTACCTGATCAGAACAAGGGGAACAACGCCtTACTttACATCCcTGACTACcGATtGG
CTGAGGGATTGTCTAATAGAATGCCACACAACCAGTCACAGGATTTCAGCACCACCAGCTCTCACAACAGCTCAGAAAGGAGTGGCAGTCT
CTCAGGTGGGAAAGGTGGAAAAAAGAAGAAAACTAAAAATTCTTCGAAAGCGCAGAAAAACAACGGATCCACTTGGGCTAATGTCCCTCTA
CCTCCTCCTCCCGTCCAGCCCCTTCCTGGTACAGAGCTGGGCCACTATGCTGCGGAACAAGAAAATGGCTATGACAGTGACAGCTGGTGCC
CACCATTACCGGTGCAAACATACTTGCATCAGGGTATGGAAGATGAGCTGGAAGAAGACGAAGATCGGGTCCCAACACCTCCTGTTCGCGG

No.13(for)+ No.12(rev)

CGTGGCCTCTTCACCAGCTATCTCTTTTGGACAGCAGTCCACTGCCACTCTTACTCCATCCCCACGGGAAGAGATGCAACCCATGCTGCAA
GCTCACTTGGATGAGTTGACAAGGGCCTATCAGTTTGATATAGCAAAACAAACATGGCACATTCAAAGCAATACCCCACCTCCACAACCCC
CAGCTCCGCCATTAGGTTACGTGTCCGGAGCCTTGATTTCTGATTTGGAGACAGATGTTCCAGATGAGGATGCTGATGATGAAGAGGAACC
ATTAGAAATTCCCAGGCCCCTCAGAGCACTAGACCAGACACCTGGATCCAGTATGGACAATCTAGACAGCTCTGTCACAGGAAAAGCCTTT
AGCTCCTCTCAAAGGCAGCGGCCCACCAGCCCATTTTCTACGGACAGTAACACCAGTGCTGCCCAGAATCAAAGCCAGAGGCCTCGGCCCA
CGAAAAAACATAAGGGAGGACAGATGGACCCACAGCCAGTGTTACCTCATCGGAGGGAAGGGATGCCAGATGATCTTCCCCCACCACCAGA
TCCCCCACCAGGTCAGGGTTTAAGGCAGCAGATAGGCCTGAGCCAACACTCTGGTAACGTGGAAAACTCAACAGAGAGAAAAGGAAGCTCT
CTAGAGAGACAACAAGCAGCCAACTTAGAAGACACAAAGAGCTCATTGGATTGTCCAGCTAAAACTGTTCTAGAGTGGCAGCGTCAAACCC

No.15

AGGACTGGATAAACTCTACAGAACGCCAAGAAGAAACTCGGAAAGCCCCACACAAACAAGGGGTTGGATCAGAGGAGTCCCTGGTGCCCTA
CAGCAAGCCCAGCTTCCCGTCTCCTGGCGGGCACAGCTCATCAGGAACATCTTCCTCTAAAGGGTCCACTGGCCCTCGGAAGGCTGACGTG
CTGAGAGGCAGCCACCAGAGGAATGCCAATGACCTTCTCGACATTGGATATGTGGGCTCAAATAGTCAAGGACAGTTTACAGAGTAACTGA
GAGGGGACATACAAAGATGCTCTGAGGACCATCAGGTCCAAACTCATGGAAGTGATAACATTTAACAGTGCAATGAACAATTTCTTTATTT
ATGTACTATTAAAAAAAAAACTGTAAATGCAATGTAAAGACACACAGCCACACATATTCCACAGTTATTTTCTTGTGTTCTTCTCTTACTTA
CACCACCCACCTTAACTCTTTCTTGTCAGGAGTATATAAGAAAAAGAAAGAAAGCAAAACTCACCCTCCAGGAAGAAAAGGATTTTTCCTC
TGTATATAATTTCTTTTGTGCATTGCTATGCAAGCTCACTCTTTTTAACTCTGTTCATATTATTGTCTGTTCTGATTGGTCTGTTGTACTA
TATGTGAATTAATGGGCTGTGGTGCCATATATTAACTTTTAATTGTGTAACTTTTTATGTTTAAATTTTGCACTGCGATTTTATTTGGTGA

No.16

TAAGCACAAATCTCTACTCCTCATGACATGAAGAAAAGATTGAATGTGAAGGGAGTTTTCTGTACTGTAAGTTAGGTTGGATAATGCTGGT
GTAACCAATCCAGTTAGATGGTTTTCAGTTGGGGGTGTAGAAATAGGAAGATCGAAGGAATGATGGTGTTGGCAAAGTCTTCTTGAAACAA
CAGATATTGAGACAATTTTAAGAAGCAGAAAGATGGATACTATTGACTAAAGCAGGGGTCAAAAGAAGGGGGTTTAAGTCTAGACAGAGTA
TGTAATAAAGTATGGTGGTAGCAAAGATGTACTAACTTGCTTTAAAAATATTATTAAAGTTTTATTTAGAATGAACTTTACCTGCCATTGT
AATTAACCCATCTTAGAATTACAATGAGCAAAATAAAAATCAAGGTGTTTCAAGAGATCTGTATTTATATTACAGCTTTTTAAAAACAGCA

No.14

GTTTTGAATAATCATAATTAATCTCTCCAACTCATTAAGTCAGACTATAACATGGAGTTCCCCAAGGAAACAAACAAAAAAAATGAACTCC
AGAAACTCTAGCAAATAGTTAAAGAAGCAACTTATTCTTCGAGCATACTCGGCTACTCTTGAATATGAGCTCAATTGCAATATCTTGTTCA
TCTTTCTAATACATGTACAGTACATATTAGAGGATGTAGGCACATCACTTAAGTTCATTTCTTTTTAAAAAAAAAAAGCAAACTAATGCAT
TCTGCATACAATTTTGTGTTTCTTTTGGTTAAAATGTGAAATCGATGCCACCAGATGTATAGCCACATTATAAGTGCTTAAAGGCAGTGCT
CACAGTATGTTCAGTTTCCAGTTAGTAGATTTATTGGACTCAATATTTTCTGGTACATTCTCAGAAATTGGCTACCAACTCAAATCTGTTT
GACCCATTTTGAATGAGTAAATCGGAAAGAACAGTCAAAAGGGAGAAAAAAAATGCATGTTTATACACTTTTTAAATAAAACCAAATCTTGT
AAGTGGACATTAATAACAGTGTCCTGCCTCATTTGTTTTCACGACTCTGAGAACAAGGAGTTGCTCAACATCCATTATGTATGCTCAAGAT
AAATGTGACTTTGAAATGTATGTTAGCTACTGTACATGTATGTAAGTCAAGTAGATGATAGCCTTCCTGAAACTACCACTAATAATATTTT
CCCATGTGCTACTTACACAAGCTTAAATTCTAGTTCAGCAAATGTTTTTCATTATTTTTTCTTACGGTTTATTCAATAGAGGGTACCTAGA
AAAATATTCAGTGTGATGGTCAACATATTGAGTGTTCAAGTAGGGGATATTTACTTTGGCAATACATACTGTGCATTATTGGTTCAGATGC
CAATTCATGAGACATTATCATAGGAAATGGATAATATAAGAACATGGCTCTAAGGGTCCTTCTCCACAGGCAAGCATAGAATTTGAGGTTC
TTCATGGGTCCCAAAACTTGCTTAGTGTTATATACTTAACTTCCTTCACTGAAGTATCAAGCAATTCTAACACACCGTTTAACAACTTTCT

ACCATCAGTGCCTGCATTTTATTGAAGTTTGATGTCTCTGGTACAAGGAAATAGAGAAAGCTGAACAGTTTCTGTTGGGAGAAAAAGTTTC
TGTAATTATTGTCAATTCTAAGTTGTGATGAATATAACTTTTCATCTATACAGTTCCTTACTCATGGTCTCTATTTAAAAGCATAGGATCT
TTTAAGATATACTTTTAGCTGAGTCTTCACTGTATGTTCATGAATAATATAATTTAGTGAAAAACTTTAGTACATCATATGTATGGTAACT
GACATTTGCAGATTATTATCAAACCAAATCATAAGAAATAATATCTTCAGATCACATGATGGCTTACCGAACAATTTCAGGTTATAGCATG
ATGTCTTTCTGATTGTCTTCAGTATATGATCATAAGCAATTATATTGTGTCCATACAGTTTACATACTGTGATTCCCGCAAAATCTTGCAT
CTGATATCTACCATGAAGATGAAATAATATGCTATGAATTTCAGGCATAGACATACTATCAGCAATAGAAAGCCTTCATCATCTAGATATC
TGGATAAATCCATTCATCGTCAACCATCAGATTGTTAACTGGGGAACACGTACCTCCCTCACAAATGCTTTCTTTTCGTCTTGTAGTTGAT
TGAATGTGATAGGAAATCTGCTGAGTTGACTCTTCTGTGGGCCTGAGCAGAAACATGCATAAGCCATCAGAAGGAAGTCTGGATGGGAAGG
AAACCTTGTCTTTCACTCGTAGGCAATACTGCTGTGTAGTCACGGCAGACCCAAGCACAGCTTCTTTCAGAACATGCAAAGTCTCCTTATG
AAATATACACATCAGTGCCCAAACTCAGGTGTACCAGTGTAAAGCGTAGGGTTTGCCTACCCTTTTTTCATATCATTTTACATAAGCAAGT
TAGGAGACTACAGGGCTATGAAAATAAATGTTTCCTTAACAGGATTTGGAGCAAGTTCACGGGAGTACATAGGACAGTCGAGCCCAGTGTA
AACACTACCAAAGGAAAGGACAGCCTTGCAGAAGCGTGAGCTCCGACTAACAGAGCGGCTTAAGTATGCTCTGTGATACAGCTTAATGCAC
TCCCTATTCAAATCCCACTGTCAATGGGATTTAAGTTGCCAGTTTATATGTGATTTCATGGAACACAACCAAGTTCTACTTAATACCTTTA
TGCAATTTTTTTTTCACTGATGAAAAATCAATATCCACATCAATGGGTGCCCAGCTGCGGATCACCTCTCCTAAAAAGCGATCTTGTCAGT
ATTGAAGAAATGTAGAACCATAATTTTCACTACTAATATCATTTTGCTCAACTACTTCTGTATATTGTGGAAAAAGGGGTGGGGCCGGGGG
AATGTGTGTTTCCACATTTTCTGAGAGGTGAAAATATTTTTGTGTTGCTTTGTCCTAATGGGTACTTCAAATGAAACACCATACAATTTTA
AATGACATGCAATTAACGAATATTTTTTACTCTCATAACATTAAAAGCATTTGTTGTG

Positions of primers are marked in gray.

## Supplementary figure 8. Mouse *Robo2a* variant 4.

No.4

CTCTGTAGCTGACAGGCTGGGAGGAGAGGAAAGAAAACGAAGGAGGAGTGCTCTGGGCGGGGAGTGGAGAAACAAAAATCTGGGCGCATTT
AGAGCATCAGGTCCTGGGGGTGTTGTTCAGCTGCAGCGCACTCTTAATTCAGGCGGACCCTGCCTGAGTTGAGTTTCTCAGCAGATTGTCA
GTCCCTGTTGGAGCAGGCTGCCATCTAGACCTGACTCCTGAGTTTAAAATGCAATGGCCAGAAGACAAGAAAGTATCTGTGGAAGGCCGTG
GACGTGGACGCCAGGACTACTGATGCTGATCATCCTTGGCATTCATCAGGGGAGTGGGCAAGGTCAAGGATCTCGTCTTCGTCAAGAAGAC
TTTCCCCCCAGGATCGTAGAACACCCTTCTGATGTCATCGTCTCCAAGGGAGAGCCCACCACTCTGAACTGTAAAGCAGAGGGCCGACCCA
CCCCCACCATTGAATGGTACAAGGATGGTGAGAGGGTGGAGACAGACAAGGATGATCCCAGGTCCCACAGAATGCTTCTGCCCAGCGGATC
TTTATTCTTTTTGCGAATTGTTCATGGGCGCAGAAGTAAACCGGACGAAGGGAGTTATGTTTGTGTTGCAAGGAACTATCTTGGTGAAGCA

No.5

GTGAGTCGAAATGCATCTCTGGAAGTGGCATTATTGCGAGATGACTTCCGGCAAAACCCCACAGATGTGGTAGTCGCAGCTGGAGAGCCTG
CAATCTTGGAGTGCCAGCCACCACGGGGACACCCAGAACCAACCATCTACTGGAAAAAGGACAAAGTCCGAATTGATGACAAGGAAGAGAG
AATAAGTATCCGTGGTGGGAAGCTGATGATCTCTAATACTAGGAAAAGCGATGCTGGCATGTACACCTGTGTGGGAACCAATATGGTGGGA
GAAAGGGACAGCGACCCTGCAGAGCTCACTGTCTTTGAACGACCCACATTTCTCAGGAGGCCAATTAACCAGGTGGTGCTAGAGGAAGAAG
CTGTAGAGTTCCGTTGTCAGGTCCAAGGAGATCCCCAGCCAACGGTGAGGTGGAAAAAAGATGATGCAGACTTGCCGAGAGGAAGGTACGA
TATCAAAGATGACTACACGCTGAGAATTAAAAAGGCCATGAGTACTGATGAAGGTACCTATGTGTGTATTGCTGAGAATCGGGTGGGAAAA

No.6

GTGGAAGCCTCTGCTACCCTCACTGTCCGAGTTCGCCCTGTTGCTCCTCCACAGTTTGTGGTTAGGCCAAGAGATCAGATCGTTGCTCAAG
GCCGAACAGTGACATTCCCCTGTGAAACTAAAGGAAACCCACAGCCAGCTGTTTTTTGGCAGAAAGAAGGCAGCCAGAACCTACTTTTCCC
GAATCAACCTCAGCAGCCCAACAGCCGATGTTCAGTGTCGCCCACGGGGGACCTCACCATCACCAACATCCAGCGTTCAGATGCGGGTTAC
TACATCTGCCAGGCCCTAACCGTGGCAGGAAGCATTTTAGCTAAAGCACAGTTGGAAGTTACTGACGTTTTGACAGATAGACCTCCACCCA
TAATCTTGCAAGGACCAATAAACCAAACACTTGCAGTAGACGGTACAGCATTGTTGAAGTGTAAAGCCACTGGTGAGCCTCTGCCTGTAAT
TAGCTGGCTAAAGGAGGGCTTTACTTTTCTGGGGAGAGATCCAAGAGCCACGATCCAAGACCAAGGAACACTGCAGATTAAGAATTTACGG

No.1(rev)+No.7(for)

ATATCTGATACTGGCACTTATACTTGTGTGGCTACAAGTTCCAGTGGAGAGACTTCCTGGAGTGCAGTGCTGGATGTAACAGAATCTGGAG
CAACAATCAGTAAAAATTATGATATGAATGACCTCCCGGGACCACCATCCAAACCTCAGGTCACTGATGTTTCTAAGAACAGTGTCACCTT

No.9

ATCCTGGCAGCCAGGTACACCTGGCGTTCTTCCTGCAAGCGCGTATATCATTGAGGCTTTCAGCCAATCGGTGAGCAATAGCTGGCAGACA
GTGGCAAACCATGTTAAGACAACTCTGTATACAGTAAGGGGGCTGAGGCCCAACACAATCTACTTGTTTATGGTCAGAGCGATCAACCCAC
AAGGTCTCAGTGATCCAAGTCCTATGTCGGATCCTGTACGCACACAAGATATCAGCCCCCCAGCACAAGGAGTGGACCACAGACAGGTACA
GAAGGAATTAGGTGATGTCGTTGTTCGTCTCCATAATCCAGTTGTCCTGACACCTACAACTGTTCAAGTCACATGGACGGTGGACCGACAA
CCCCAGTTTATTCAGGGCTACAGAGTGATGTACCGTCAGACTTCGGGACTACAAGCCTCAACTGTGTGGCAGAATCTAGACGCCAAAGTCC
CGACTGAGAGGAGTGCTGTCCTTGTGAATTTGAAAAAGGGGGTGACTTATGAAATTAAAGTCCGGCCGTATTTTAACGAGTTCCAAGGAAT
GGACAGTGAATCGAAAACAGTCCGAACCACTGAGGAAGCCCCAAGTGCCCCTCCCCAGTCTGTCACTGTGCTGACAGTTGGAAGTCACAAC
AGCACAAGCATCAGTGTTTCCTGGGATCCTCCACCAGCCGACCACCAGAATGGAATTATTCAGGAATATAAGATCTGGTGTCTGGGAAACG

No.10

AAACGCGATTCCATATCAATAAAACGGTGGATGCAGCCATTCGCTCTGTAGTAATAGGTGGCTTGTTCCCTGGAATTCAGTACCGGGTAGA
AGTGGCAGCTAGCACAAGTGCAGGGGTTGGAGTAAAAAGTGAACCACAGCCGATAATAATTGGGGGACGTAACGAAGTTGTCATTACTGAA
AACAATAACAGCATCACTGAGCAAATCACGGATGTCGTGAAGCAACCGGCATTTATAGCTGGCATTGGTGGTGCCTGCTGGGTAATTCTGA

No.11(for)+ No.8(rev)

TGGGTTTTAGCATCTGGTTGTACTGGAGAAGAAAGAAGAGAAAGGGACTCAGCAATTATGCTGTAACATTTCAAAGAGGAGATGGAGGACT
AATGAGCAATGGGAGCCGTCCAGGTCTTCTAAATGCTGGCGATCCCAATTACCCATGGCTTGCTGATTCTTGGCCAGCCACGAGTTTGCCA
GTGAACAATAGCAATAGTGGCCCAAATGAAATTGGAAATTTTGGGCGTGGAGATGTGCTGCCTCCGGTGCCAGGCCAAGGGGATAAAACAG

CGACCATGCTCTCGGATGGAGCCATTTATAGCAGCATTGACTTCACTACCAAAACCACTTACAACAGTTCCAGCCAAATAACACAGGCCAC
CCCATATGCCACTACACAAATCCTGCATTCAAACAGCATCCACGAACTGGCAGTTGATCTTCCTGATCCACAGTGGAAAAGCTCAGTTCAA
CAGAAGACAGACCTCATGGGATTTGGTTATTCGCTACCTGATCAGAACAAGGGGAACAACGCCtTACTttACATCCCTGACTACCGATtGG
CTGAGGGATTGTCTAATAGAATGCCACACAACCAGTCACAGGATTTCAGCACCACCAGCTCTCACAACAGCTCAGAAAGGAGTGGCAGTCT
CTCAGGTGGGAAAGGTGGAAAAAAGAAGAAAACTAAAAATTCTTCGAAAGCGCAGAAAAACAACGGATCCACTTGGGCTAATGTCCCTCTA
CCTCCTCCTCCCGTCCAGCCCCTTCCTGGTACAGAGCTGGGCCACTATGCTGCGGAACAAGAAAATGGCTATGACAGTGACAGCTGGTGCC
CACCATTACCGGTGCAAACATACTTGCATCAGGGTATGGAAGATGAGCTGGAAGAAGACGAAGATCGGGTCCCAACACCTCCTGTTCGCGG

No.13(for)+ No.12(rev)

CGTGGCCTCTTCACCAGCTATCTCTTTTGGACAGCAGTCCACTGCCACTCTTACTCCATCCCCACGGGAAGAGATGCAACCCATGCTGCAA
GCTCACTTGGATGAGTTGACAAGGGCCTATCAGTTTGATATAGCAAAACAAACATGGCACATTCAAAGCAATACCCCACCTCCACAACCCC
CAGCTCCGCCATTAGGTTACGTGTCCGGAGCCTTGATTTCTGATTTGGAGACAGATGTTCCAGATGAGGATGCTGATGATGAAGAGGAACC
ATTAGAAATTCCCAGGCCCCTCAGAGCACTAGACCAGACACCTGGATCCAGTATGGACAATCTAGACAGCTCTGTCACAGGAAAAGCCTTT
AGCTCCTCTCAAAGGCAGCGGCCCACCAGCCCATTTTCTACGGACAGTAACACCAGTGCTGCCCAGAATCAAAGCCAGAGGCCTCGGCCCA

No.15

CGAAAAAACATAAGGGAGGACAGATGGACCCACAGCCAGTGTTACCTCATCGGAGGGAAGGGATGCCAGATGAGGAGTCCCTGGTGCCCTA
CAGCAAGCCCAGCTTCCCGTCTCCTGGCGGGCACAGCTCATCAGGAACATCTTCCTCTAAAGGGTCCACTGGCCCTCGGAAGGCTGACGTG
CTGAGAGGCAGCCACCAGAGGAATGCCAATGACCTTCTCGACATTGGATATGTGGGCTCAAATAGTCAAGGACAGTTTACAGAGTAACTGA
GAGGGGACATACAAAGATGCTCTGAGGACCATCAGGTCCAAACTCATGGAAGTGATAACATTTAACAGTGCAATGAACAATTTCTTTATTT
ATGTACTATTAAAAAAAAACTGTAAATGCAATGTAAAGACACACAGCCACACATATTCCACAGTTATTTTCTTGTGTTCTTCTCTTACTTA
CACCACCCACCTTAACTCTTTCTTGTCAGGAGTATATAAGAAAAAGAAAGAAAGCAAAACTCACCCTCCAGGAAGAAAAGGATTTTTCCTC
TGTATATAATTTCTTTTGTGCATTGCTATGCAAGCTCACTCTTTTTAACTCTGTTCATATTATTGTCTGTTCTGATTGGTCTGTTGTACTA
TATGTGAATTAATGGGCTGTGGTGCCATATATTAACTTTTAATTGTGTAACTTTTTATGTTTAAATTTTGCACTGCGATTTTATTTGGTGA

No.16

TAAGCACAAATCTCTACTCCTCATGACATGAAGAAAAGATTGAATGTGAAGGGAGTTTTCTGTACTGTAAGTTAGGTTGGATAATGCTGGT
GTAACCAATCCAGTTAGATGGTTTTCAGTTGGGGGTGTAGAAATAGGAAGATCGAAGGAATGATGGTGTTGGCAAAGTCTTCTTGAAACAA
CAGATATTGAGACAATTTTAAGAAGCAGAAAGATGGATACTATTGACTAAAGCAGGGGTCAAAAGAAGGGGGTTTAAGTCTAGACAGAGTA
TGTAATAAAGTATGGTGGTAGCAAAGATGTACTAACTTGCTTTAAAAATATTATTAAAGTTTTATTTAGAATGAACTTTACCTGCCATTGT
AATTAACCCATCTTAGAATTACAATGAGCAAAATAAAAATCAAGGTGTTTCAAGAGATCTGTATTTATATTACAGCTTTTTAAAAACAGCA

No.14

GTTTTGAATAATCATAATTAATCTCTCCAACTCATTAAGTCAGACTATAACATGGAGTTCCCCAAGGAAACAAACAAAAAAAATGAACTCC
AGAAACTCTAGCAAATAGTTAAAGAAGCAACTTATTCTTCGAGCATACTCGGCTACTCTTGAATATGAGCTCAATTGCAATATCTTGTTCA
TCTTTCTAATACATGTACAGTACATATTAGAGGATGTAGGCACATCACTTAAGTTCATTTCTTTTTAAAAAAAAAAAGCAAACTAATGCAT
TCTGCATACAATTTTGTGTTTCTTTTGGTTAAAATGTGAAATCGATGCCACCAGATGTATAGCCACATTATAAGTGCTTAAAGGCAGTGCT
CACAGTATGTTCAGTTTCCAGTTAGTAGATTTATTGGACTCAATATTTTCTGGTACATTCTCAGAAATTGGCTACCAACTCAAATCTGTTT
GACCCATTTTGAATGAGTAAATCGGAAAGAACAGTCAAAAGGGAGAAAAAAATGCATGTTTATACACTTTTTAAATAAAACCAAATCTTGT
AAGTGGACATTAATAACAGTGTCCTGCCTCATTTGTTTTCACGACTCTGAGAACAAGGAGTTGCTCAACATCCATTATGTATGCTCAAGAT
AAATGTGACTTTGAAATGTATGTTAGCTACTGTACATGTATGTAAGTCAAGTAGATGATAGCCTTCCTGAAACTACCACTAATAATATTTT
CCCATGTGCTACTTACACAAGCTTAAATTCTAGTTCAGCAAATGTTTTTCATTATTTTTTCTTACGGTTTATTCAATAGAGGGTACCTAGA
AAAATATTCAGTGTGATGGTCAACATATTGAGTGTTCAAGTAGGGGATATTTACTTTGGCAATACATACTGTGCATTATTGGTTCAGATGC
CAATTCATGAGACATTATCATAGGAAATGGATAATATAAGAACATGGCTCTAAGGGTCCTTCTCCACAGGCAAGCATAGAATTTGAGGTTC
TTCATGGGTCCCAAAACTTGCTTAGTGTTATATACTTAACTTCCTTCACTGAAGTATCAAGCAATTCTAACACACCGTTTAACAACTTTCT
ACCATCAGTGCCTGCATTTTATTGAAGTTTGATGTCTCTGGTACAAGGAAATAGAGAAAGCTGAACAGTTTCTGTTGGGAGAAAAAGTTTC
TGTAATTATTGTCAATTCTAAGTTGTGATGAATATAACTTTTCATCTATACAGTTCCTTACTCATGGTCTCTATTTAAAAGCATAGGATCT
TTTAAGATATACTTTTAGCTGAGTCTTCACTGTATGTTCATGAATAATATAATTTAGTGAAAAACTTTAGTACATCATATGTATGGTAACT

GACATTTGCAGATTATTATCAAACCAAATCATAAGAAATAATATCTTCAGATCACATGATGGCTTACCGAACAATTTCAGGTTATAGCATG
ATGTCTTTCTGATTGTCTTCAGTATATGATCATAAGCAATTATATTGTGTCCATACAGTTTACATACTGTGATTCCCGCAAAATCTTGCAT
CTGATATCTACCATGAAGATGAAATAATATGCTATGAATTTCAGGCATAGACATACTATCAGCAATAGAAAGCCTTCATCATCTAGATATC
TGGATAAATCCATTCATCGTCAACCATCAGATTGTTAACTGGGGAACACGTACCTCCCTCACAAATGCTTTCTTTTCGTCTTGTAGTTGAT
TGAATGTGATAGGAAATCTGCTGAGTTGACTCTTCTGTGGGCCTGAGCAGAAACATGCATAAGCCATCAGAAGGAAGTCTGGATGGGAAGG
AAACCTTGTCTTTCACTCGTAGGCAATACTGCTGTGTAGTCACGGCAGACCCAAGCACAGCTTCTTTCAGAACATGCAAAGTCTCCTTATG
AAATATACACATCAGTGCCCAAACTCAGGTGTACCAGTGTAAAGCGTAGGGTTTGCCTACCCTTTTTTCATATCATTTTACATAAGCAAGT
TAGGAGACTACAGGGCTATGAAAATAAATGTTTCCTTAACAGGATTTGGAGCAAGTTCACGGGAGTACATAGGACAGTCGAGCCCAGTGTA
AACACTACCAAAGGAAAGGACAGCCTTGCAGAAGCGTGAGCTCCGACTAACAGAGCGGCTTAAGTATGCTCTGTGATACAGCTTAATGCAC
TCCCTATTCAAATCCCACTGTCAATGGGATTTAAGTTGCCAGTTTATATGTGATTTCATGGAACACAACCAAGTTCTACTTAATACCTTTA
TGCAATTTTTTTTTCACTGATGAAAAATCAATATCCACATCAATGGGTGCCCAGCTGCGGATCACCTCTCCTAAAAAGCGATCTTGTCAGT
ATTGAAGAAATGTAGAACCATAATTTTCACTACTAATATCATTTTGCTCAACTACTTCTGTATATTGTGGAAAAAGGGGTGGGGCCGGGGG
AATGTGTGTTTCCACATTTTCTGAGAGGTGAAAATATTTTTGTGTTGCTTTGTCCTAATGGGTACTTCAAATGAAACACCATACAATTTTA
AATGACATGCAATTAACGAATATTTTTTACTCTCATAACATTAAAAGCATTTGTTGTG

Positions of primers are marked in gray.

## Supplementary figure 9. Mouse *Robo2b* variant 1.

No.3

ATGAATCCTCTGATGTTTACACTATTATTGCTCTTTGGATTTCTCTGCATTCAGATTGATGGATCTCGTCTTCGTCAAGAAGACTTTCCCC
CCAGGATCGTAGAACACCCTTCTGATGTCATCGTCTCCAAGGGAGAGCCCACCACTCTGAACTGTAAAGCAGAGGGCCGACCCACCCCCAC
CATTGAATGGTACAAGGATGGTGAGAGGGTGGAGACAGACAAGGATGATCCCAGGTCCCACAGAATGCTTCTGCCCAGCGGATCTTTATTC
TTTTTGCGAATTGTTCATGGGCGCAGAAGTAAACCGGACGAAGGGAGTTACGTTTGTGTTGCAAGGAACTATCTTGGTGAAGCAGTGAGTC

No.5

GAAATGCATCTCTGGAAGTGGCATTATTGCGAGATGACTTCCGGCAAAACCCCACAGATGTGGTAGTCGCAGCTGGAGAGCCTGCAATCTT
GGAGTGCCAGCCACCACGGGGACACCCAGAACCAACCATCTACTGGAAAAAGGACAAAGTCCGAATTGATGACAAGGAAGAGAGAATAAGT
ATCCGTGGTGGGAAGCTGATGATCTCTAATACTAGGAAAAGCGATGCTGGCATGTACACCTGTGTGGGAACCAATATGGTGGGAGAAAGGG
ACAGCGACCCTGCAGAGCTCACTGTCTTTGAACGACCCACATTTCTCAGGAGGCCAATTAACCAGGTGGTGCTAGAGGAAGAAGCTGTAGA
ATTCCGTTGTCAGGTCCAAGGAGATCCCCAGCCAACGGTGAGGTGGAAAAAAGATGATGCAGACTTGCCGAGAGGAAGGTATGATATCAAA
GATGACTACACGCTGAGAATTAAAAAGGCCATGAGTACTGATGAAGGTACCTATGTGTGTATTGCTGAGAATCGGGTGGGAAAAGTGGAAG

No.6

CCTCTGCTACCCTCACTGTCCGAGCTCCTCCACAGTTTGTGGTTAGGCCAAGAGATCAGATCGTTGCTCAAGGCCGAACAGTGACATTCCC
CTGTGAAACTAAAGGAAACCCACAGCCAGCTGTTTTTTGGCAGAAAGAAGGCAGCCAGAACCTACTTTTCCCGAATCAACCTCAGCAGCCC
AACAGCCGATGTTCAGTGTCGCCCACGGGGGACCTCACCATCACCAACATCCAGCGTTCAGATGCGGGTTACTACATCTGCCAGGCCCTAA
CCGTGGCAGGAAGCATTTTAGCTAAAGCACAGTTGGAAGTTACTGACGTTTTGACAGATAGACCTCCACCCATAATCTTGCAAGGACCAAT
AAACCAAACACTTGCAGTAGACGGTACAGCATTGTTGAAGTGTAAAGCCACTGGTGAGCCTCTGCCTGTAATTAGCTGGCTAAAGGAGGGC

No.1 (rev) + No.7 (for)

TTTACTTTTCTGGGGAGAGATCCAAGAGCCACGATCCAAGACCAAGGAACACTGCAGATTAAGAATTTACGGATATCTGATACTGGCACTT
ATACTTGTGTGGCTACAAGTTCCAGTGGAGAGACTTCCTGGAGTGCAGTGCTGGATGTAACAGAATCTGGAGCAACAATCAGTAAAAATTA
TGATATGAATGACCTCCCGGGACCACCATCCAAACCTCAGGTCACTGATGTTTCTAAGAACAGTGTCACCTTATCCTGGCAGCCAGGTACA

No.9

CCTGGCGTTCTTCCTGCAAGCGCGTATATCATTGAGGCTTTCAGCCAATCGGTGAGCAATAGCTGGCAGACAGTGGCAAACCATGTTAAGA
CAACTCTGTATACAGTAAGGGGGCTGAGGCCCAACACAATCTACTTGTTTATGGTCAGAGCGATCAACCCACAAGGTCTCAGTGATCCAAG
TCCTATGTCGGATCCTGTACGCACACAAGATATCAGCCCCCCAGCACAAGGAGTGGACCACAGACAGGTGCAGAAGGAATTAGGTGATGTC
GTTGTTCGTCTCCATAATCCAGTTGTCCTGACACCTACAACTGTTCAAGTCACATGGACGGTGGACCGACAACCCCAGTTTATTCAGGGCT
ACAGAGTGATGTACCGTCAGACTTCGGGACTACAAGCCTCAACTGTGTGGCAGAATCTAGACGCCAAAGTCCCGACTGAGAGGAGTGCTGT
CCTTGTGAATTTGAAAAAGGGGGTGACTTATGAAATTAAAGTCCGGCCGTATTTTAACGAGTTCCAAGGAATGGACAGTGAATCGAAAACA
GTCCGAACCACTGAGGAAGCCCCAAGTGCCCCTCCCCAGTCTGTCACTGTGCTGACAGTTGGAAGTCACAACAGCACAAGCATCAGTGTTT
CCTGGGATCCTCCACCAGCCGACCACCAGAATGGAATTATTCAGGAATATAAGATCTGGTGTCTGGGAAACGAAACGCGATTCCATATCAA

No.10

TAAAACGGTGGATGCAGCCATTCGCTCTGTAGTAATAGGTGGCTTGTTCCCTGGAATTCAGTACCGGGTAGAAGTGGCAGCTAGCACAAGT
GCAGGGGTTGGAGTAAAAAGTGAACCACAGCCGATAATAATTGGGGGACGTAACGAAGTTGTCATTACTGAAAACAATAACAGCATCACTG

No.11 (for) + No.8 (rev)

AGCAAATCACGGATGTCGTGAAGCAACCGGCATTTATAGCTGGCATTGGTGGTGCCTGCTGGGTAATTCTGATGGGTTTTAGCATCTGGTT
GTACTGGAGAAGAAAGAAGAGAAAGGGACTCAGCAATTATGCTGTAACATTTCAAAGAGGAGATGGAGGACTAATGAGCAATGGGAGCCGT
CCAGGTCTTCTAAATGCTGGCGATCCCAATTACCCATGGCTTGCTGATTCTTGGCCAGCCACGAGTTTGCCAGTGAACAATAGCAATAGTG
GCCCAAATGAAATTGGAAATTTTGGGCGTGGAGATGTGCTGCCTCCGGTGCCAGGCCAAGGGGATAAAACAGCGACCATGCTCTCGGATGG
AGCCATTTATAGCAGCATTGACTTCACTACCAAAACCACTTACAACAGTTCCAGCCAAATAACACAGGCCACCCCATATGCCACTACACAA
ATCCTGCATTCAAACAGCATCCACGAACTGGCAGTTGATCTTCCTGATCCACAGTGGAAAAGCTCAGTTCAACAGAAGACAGACCTCATGG
GATTTGGTTATTCGCTACCTGATCAGAACAAGGGGAACAACGCCtTACTttACATCCcTGACTACcGATtGGCTGAGGGATTGTCTAATAG

AATGCCACACAACCAGTCACAGGATTTCAGCACCACCAGCTCTCACAACAGCTCAGAAAGGAGTGGCAGTCTCTCAGGTGGGAAAGGTGGA
AAAAAGAAGAAAACTAAAAATTCTTCGAAAGCGCAGAAAAACAACGGATCCACTTGGGCTAATGTCCCTCTACCTCCTCCTCCCGTCCAGC
CCCTTCCTGGTACAGAGCTGGGCCACTATGCTGCGGAACAAGAAAATGGCTATGACAGTGACAGCTGGTGCCCACCATTACCGGTGCAAAC
ATACTTGCATCAGGGTATGGAAGATGAGCTGGAAGAAGACGAAGATCGGGTCCCAACACCTCCTGTTCGCGGCGTGGCCTCTTCACCAGCT

No.13 (for) + No.12 (rev)

ATCTCTTTTGGACAGCAGTCCACTGCCACTCTTACTCCATCCCCACGGGAAGAGATGCAACCCATGCTGCAAGCTCACTTGGATGAGTTGA
CAAGGGCCTATCAGTTTGATATAGCAAAACAAACATGGCACATTCAAAGCAATACCCCACCTCCACAACCCCCAGCTCCGCCATTAGGTTA
CGTGTCCGGAGCCTTGATTTCTGATTTGGAGACAGATGTTCCAGATGAGGATGCTGATGATGAAGAGGAACCATTAGAAATTCCCAGGCCC
CTCAGAGCACTAGACCAGACACCTGGATCCAGTATGGACAATCTAGACAGCTCTGTCACAGGAAAAGCCTTTAGCTCCTCTCAAAGGCAGC
GGCCCACCAGCCCATTTTCTACGGACAGTAACACCAGTGCTGCCCAGAATCAAAGCCAGAGGCCTCGGCCCACGAAAAAACATAAGGGAGG
ACGGATGGACCCACAGCCAGTGTTACCTCATCGGAGGGAAGGGATGCCAGATGATCTTCCCCCACCACCAGATCCCCCACCAGGTCAGGGT
TTAAGGCAGCAGATAGGCCTGAGCCAACACTCTGGTAACGTGGAAAACTCAACAGAGAGAAAAGGAAGCTCTCTAGAGAGACAACAAGCAG
CCAACTTAGAAGACACAAAGAGCTCATTGGATTGTCCAGCTAAAACTGTTCTAGAGTGGCAGCGTCAAACCCAGGACTGGATAAACTCTAC

No.15

AGAACGCCAAGAAGAAACTCGGAAAGCCCCACACAAACAAGGGGTTGGATCAGAGGAGTCCCTGGTGCCCTACAGCAAGCCCAGCTTCCCG
TCTCCTGGCGGGCACAGCTCATCAGGAACATCTTCCTCTAAAGGGTCCACTGGCCCTCGGAAGGCTGACGTGCTGAGAGGCAGCCACCAGA
GGAATGCCAATGACCTTCTCGACATTGGATATGTGGGCTCAAATAGTCAAGGACAGTTTACAGAGTAACTGAGAGGGGACATACAAAGATG
CTCTGAGGACCATCAGGTCCAAACTCATGGAAGTGATAACATTTAACAGTGCAATGAACAATTTCTTTATTTATGTACTATTAAAAAAAAA
CTGTAAATGCAATGTAAAGACACACAGCCACACATATTCCACAGTTATTTTCTTGTGTTCTTCTCTTACTTACACCACCCACCTTAACTCT
TTCTTGTCAGGAGTATATAAGAAAAAGAAAGAAAGCAAAACTCACCCTCCAGGAAGAAAAGGATTTTTCCTCTGTATATAATTTCTTTTGT
GCATTGCTATGCAAGCTCACTCTTTTTAACTCTGTTCATATTATTGTCTGTTCTGATTGGTCTGTTGTACTATATGTGAATTAATGGGCTG
TGGTGCCATATATTAACTTTTAATTGTGTAACTTTTTATGTTTAAATTTTGCACTGCGATTTTATTTGGTGATAAGCACAAATCTCTACTC

No.16

CTCATGACATGAAGAAAAGATTGAATGTGAAGGGAGTTTTCTGTACTGTAAGTTAGGTTGGATAATGCTGGTGTAACCAATCCAGTTAGAT
GGTTTTCAGTTGGGGGTGTAGAAATAGGAAGATCGAAGGAATGATGGTGTTGGCAAAGTCTTCTTGAAACAACAGATATTGAGACAATTTT
AAGAAGCAGAAAGATGGATACTATTGACTAAAGCAGGGGTCAAAAGAAGGGGGTTTAAGTCTAGACAGAGTATGTAATAAAGTATGGTGGT
AGCAAAGATGTACTAACTTGCTTTAAAAATATTATTAAAGTTTTATTTAGAATGAACTTTACCTGCCATTGTAATTAACCCATCTTAGAAT
TACAATGAGCAAAATAAAAATCAAGGTGTTTCAAGAGATCTGTATTTATATTACAGCTTTTTAAAAACAGCAGTTTTGAATAATCATAATT

No.14

AATCTCTCCAACTCATTAAGTCAGACTATAACATGGAGTTCCCCAAGGAAACAAACAAAAAAAATGAACTCCAGAAACTCTAGCAAATAGT
TAAAGAAGCAACTTATTCTTCGAGCATACTCGGCTACTCTTGAATATGAGCTCAATTGCAATATCTTGTTCATCTTTCTAATACATGTACA
GTACATATTAGAGGATGTAGGCACATCACTTAAGTTCATTTCTTTTTAAAAAAAAAAAGCAAACTAATGCATTCTGCATACAATTTTGTGT
TTCTTTTGGTTAAAATGTGAAATCGATGCCACCAGATGTATAGCCACATTATAAGTGCTTAAAGGCAGTGCTCACAGTATGTTCAGTTTCC
AGTTAGTAGATTTATTGGACTCAATATTTTCTGGTACATTCTCAGAAATTGGCTACCAACTCAAATCTGTTTGACCCATTTTGAATGAGTA
AATCGGAAAGAACAGTCAAAAGGGAGAAAAAAATGCATGTTTATACACTTTTTAAATAAAACCAAATCTTGTAAGTGGACATTAATAACAG
TGTCCTGCCTCATTTGTTTTCACGACTCTGAGAACAAGGAGTTGCTCAACATCCATTATGTATGCTCAAGATAAATGTGACTTTGAAATGT
ATGTTAGCTACTGTACATGTATGTAAGTCAAGTAGATGATAGCCTTCCTGAAACTACCACTAATAATATTTTCCCATGTGCTACTTACACA
AGCTTAAATTCTAGTTCAGCAAATGTTTTTCATTATTTTTTCTTACGGTTTATTCAATAGAGGGTACCTAGAAAAATATTCAGTGTGATGG
TCAACATATTGAGTGTTCAAGTAGGGGATATTTACTTTGGCAATACATACTGTGCATTATTGGTTCAGATGCCAATTCATGAGACATTATC
ATAGGAAATGGATAATATAAGAACATGGCTCTAAGGGTCCTTCTCCACAGGCAAGCATAGAATTTGAGGTTCTTCATGGGTCCCAAAACTT
GCTTAGTGTTATATACTTAACTTCCTTCACTGAAGTATCAAGCAATTCTAACACACCGTTTAACAACTTTCTACCATCAGTGCCTGCATTT
TATTGAAGTTTGATGTCTCTGGTACAAGGAAATAGAGAAAGCTGAACAGTTTCTGTTGGGAGAAAAAGTTTCTGTAATTATTGTCAATTCT
AAGTTGTGATGAATATAACTTTTCATCTATACAGTTCCTTACTCATGGTCTCTATTTAAAAGCATAGGATCTTTTAAGATATACTTTTAGC
TGAGTCTTCACTGTATGTTCATGAATAATATAATTTAGTGAAAAACTTTAGTACATCATATGTATGGTAACTGACATTTGCAGATTATTAT

CAAACCAAATCATAAGAAATAATATCTTCAGATCACATGATGGCTTACCGAACAATTTCAGGTTATAGCATGATGTCTTTCTGATTGTCTT
CAGTATATGATCATAAGCAATTATATTGTGTCCATACAGTTTACATACTGTGATTCCCGCAAAATCTTGCATCTGATATCTACCATGAAGA
TGAAATAATATGCTATGAATTTCAGGCATAGACATACTATCAGCAATAGAAAGCCTTCATCATCTAGATATCTGGATAAATCCATTCATCG
TCAACCATCAGATTGTTAACTGGGGAACACGTACCTCCCTCACAAATGCTTTCTTTTCGTCTTGTAGTTGATTGAATGTGATAGGAAATCT
GCTGAGTTGACTCTTCTGTGGGCCTGAGCAGAAACATGCATAAGCCATCAGAAGGAAGTCTGGATGGGAAGGAAACCTTGTCTTTCACTCG
TAGGCAATACTGCTGTGTAGTCACGGCAGACCCAAGCACAGCTTCTTTCAGAACATGCAAAGTCTCCTTATGAAATATACACATCAGTGCC
CAAACTCAGGTGTACCAGTGTAAAGCGTAGGGTTTGCCTACCCTTTTTTCATATCATTTTACATAAGCAAGTTAGGAGACTACAGGGCTAT
GAAAATAAATGTTTCCTTAACAGGATTTGGAGCAAGTTCACGGGAGTACATAGGACAGTCGAGCCCAGTGTAAACACTACCAAAGGAAAGG
ACAGCCTTGCAGAAGCGTGAGCTCCGACTAACAGAGCGGCTTAAGTATGCTCTGTGATACAGCTTAATGCACTCCCTATTCAAATCCCACT
GTCAATGGGATTTAAGTTGCCAGTTTATATGTGATTTCATGGAACACAACCAAGTTCTACTTAATACCTTTATGCAATTTTTTTTTCACTG
ATGAAAAATCAATATCCACATCAATGGGTGCCCAGCTGCGGATCACCTCTCCTAAAAAGCGATCTTGTCAGTATTGAAGAAATGTAGAACC
ATAATTTTCACTACTAATATCATTTTGCTCAACTACTTCTGTATATTGTGGAAAAAGGGGTGGGGCCGGGGGAATGTGTGTTTCCACATTT
TCTGAGAGGTGAAAATATTTTTGTGTTGCTTTGTCCTAATGGGTACTTCAAATGAAACACCATACAATTTTAAATGACATGCAATTAACGA
ATATTTTTTACTCTCATAACATTAAAAGCATTTGTTGTG

Positions of primers are marked in gray.

## Supplementary figure 10. Mouse *Robo2b* variant 2.

No.3

ATGAATCCTCTGATGTTTACACTATTATTGCTCTTTGGATTTCTCTGCATTCAGATTGATGGATCTCGTCTTCGTCAAGAAGACTTTCCCC
CCAGGATCGTAGAACACCCTTCTGATGTCATCGTCTCCAAGGGAGAGCCCACCACTCTGAACTGTAAAGCAGAGGGCCGACCCACCCCCAC
CATTGAATGGTACAAGGATGGTGAGAGGGTGGAGACAGACAAGGATGATCCCAGGTCCCACAGAATGCTTCTGCCCAGCGGATCTTTATTC
TTTTTGCGAATTGTTCATGGGCGCAGAAGTAAACCGGACGAAGGGAGTTACGTTTGTGTTGCAAGGAACTATCTTGGTGAAGCAGTGAGTC

No.5

GAAATGCATCTCTGGAAGTGGCATTATTGCGAGATGACTTCCGGCAAAACCCCACAGATGTGGTAGTCGCAGCTGGAGAGCCTGCAATCTT
GGAGTGCCAGCCACCACGGGGACACCCAGAACCAACCATCTACTGGAAAAAGGACAAAGTCCGAATTGATGACAAGGAAGAGAGAATAAGT
ATCCGTGGTGGGAAGCTGATGATCTCTAATACTAGGAAAAGCGATGCTGGCATGTACACCTGTGTGGGAACCAATATGGTGGGAGAAAGGG
ACAGCGACCCTGCAGAGCTCACTGTCTTTGAACGACCCACATTTCTCAGGAGGCCAATTAACCAGGTGGTGCTAGAGGAAGAAGCTGTAGA
ATTCCGTTGTCAGGTCCAAGGAGATCCCCAGCCAACGGTGAGGTGGAAAAAAGATGATGCAGACTTGCCGAGAGGAAGGTATGATATCAAA
GATGACTACACGCTGAGAATTAAAAAGGCCATGAGTACTGATGAAGGTACCTATGTGTGTATTGCTGAGAATCGGGTGGGAAAAGTGGAAG

No.6

CCTCTGCTACCCTCACTGTCCGAGCTCCTCCACAGTTTGTGGTTAGGCCAAGAGATCAGATCGTTGCTCAAGGCCGAACAGTGACATTCCC
CTGTGAAACTAAAGGAAACCCACAGCCAGCTGTTTTTTGGCAGAAAGAAGGCAGCCAGAACCTACTTTTCCCGAATCAACCTCAGCAGCCC
AACAGCCGATGTTCAGTGTCGCCCACGGGGGACCTCACCATCACCAACATCCAGCGTTCAGATGCGGGTTACTACATCTGCCAGGCCCTAA
CCGTGGCAGGAAGCATTTTAGCTAAAGCACAGTTGGAAGTTACTGACGTTTTGACAGATAGACCTCCACCCATAATCTTGCAAGGACCAAT
AAACCAAACACTTGCAGTAGACGGTACAGCATTGTTGAAGTGTAAAGCCACTGGTGAGCCTCTGCCTGTAATTAGCTGGCTAAAGGAGGGC

No.1 (rev) + No.7 (for)

TTTACTTTTCTGGGGAGAGATCCAAGAGCCACGATCCAAGACCAAGGAACACTGCAGATTAAGAATTTACGGATATCTGATACTGGCACTT
ATACTTGTGTGGCTACAAGTTCCAGTGGAGAGACTTCCTGGAGTGCAGTGCTGGATGTAACAGAATCTGGAGCAACAATCAGTAAAAATTA
TGATATGAATGACCTCCCGGGACCACCATCCAAACCTCAGGTCACTGATGTTTCTAAGAACAGTGTCACCTTATCCTGGCAGCCAGGTACA

No.9

CCTGGCGTTCTTCCTGCAAGCGCGTATATCATTGAGGCTTTCAGCCAATCGGTGAGCAATAGCTGGCAGACAGTGGCAAACCATGTTAAGA
CAACTCTGTATACAGTAAGGGGGCTGAGGCCCAACACAATCTACTTGTTTATGGTCAGAGCGATCAACCCACAAGGTCTCAGTGATCCAAG
TCCTATGTCGGATCCTGTACGCACACAAGATATCAGCCCCCCAGCACAAGGAGTGGACCACAGACAGGTGCAGAAGGAATTAGGTGATGTC
GTTGTTCGTCTCCATAATCCAGTTGTCCTGACACCTACAACTGTTCAAGTCACATGGACGGTGGACCGACAACCCCAGTTTATTCAGGGCT
ACAGAGTGATGTACCGTCAGACTTCGGGACTACAAGCCTCAACTGTGTGGCAGAATCTAGACGCCAAAGTCCCGACTGAGAGGAGTGCTGT
CCTTGTGAATTTGAAAAAGGGGGTGACTTATGAAATTAAAGTCCGGCCGTATTTTAACGAGTTCCAAGGAATGGACAGTGAATCGAAAACA
GTCCGAACCACTGAGGAAGCCCCAAGTGCCCCTCCCCAGTCTGTCACTGTGCTGACAGTTGGAAGTCACAACAGCACAAGCATCAGTGTTT
CCTGGGATCCTCCACCAGCCGACCACCAGAATGGAATTATTCAGGAATATAAGATCTGGTGTCTGGGAAACGAAACGCGATTCCATATCAA

No.10

TAAAACGGTGGATGCAGCCATTCGCTCTGTAGTAATAGGTGGCTTGTTCCCTGGAATTCAGTACCGGGTAGAAGTGGCAGCTAGCACAAGT
GCAGGGGTTGGAGTAAAAAGTGAACCACAGCCGATAATAATTGGGGGACGTAACGAAGTTGTCATTACTGAAAACAATAACAGCATCACTG

No.11(for)+ No.8(rev)

AGCAAATCACGGATGTCGTGAAGCAACCGGCATTTATAGCTGGCATTGGTGGTGCCTGCTGGGTAATTCTGATGGGTTTTAGCATCTGGTT
GTACTGGAGAAGAAAGAAGAGAAAGGGACTCAGCAATTATGCTGTAACATTTCAAAGAGGAGATGGAGGACTAATGAGCAATGGGAGCCGT
CCAGGTCTTCTAAATGCTGGCGATCCCAATTACCCATGGCTTGCTGATTCTTGGCCAGCCACGAGTTTGCCAGTGAACAATAGCAATAGTG
GCCCAAATGAAATTGGAAATTTTGGGCGTGGAGATGTGCTGCCTCCGGTGCCAGGCCAAGGGGATAAAACAGCGACCATGCTCTCGGATGG
AGCCATTTATAGCAGCATTGACTTCACTACCAAAACCACTTACAACAGTTCCAGCCAAATAACACAGGCCACCCCATATGCCACTACACAA
ATCCTGCATTCAAACAGCATCCACGAACTGGCAGTTGATCTTCCTGATCCACAGTGGAAAAGCTCAGTTCAACAGAAGACAGACCTCATGG
GATTTGGTTATTCGCTACCTGATCAGAACAAGGGGAACAACGGTGGGAAAGGTGGAAAAAAGAAGAAAACTAAAAATTCTTCGAAAGCGCA

GAAAAACAACGGATCCACTTGGGCTAATGTCCCTCTACCTCCTCCTCCCGTCCAGCCCCTTCCTGGTACAGAGCTGGGCCACTATGCTGCG
GAACAAGAAAATGGCTATGACAGTGACAGCTGGTGCCCACCATTACCGGTGCAAACATACTTGCATCAGGGTATGGAAGATGAGCTGGAAG

No.13(for)+ No.12(rev)

AAGACGAAGATCGGGTCCCAACACCTCCTGTTCGCGGCGTGGCCTCTTCACCAGCTATCTCTTTTGGACAGCAGTCCACTGCCACTCTTAC
TCCATCCCCACGGGAAGAGATGCAACCCATGCTGCAAGCTCACTTGGATGAGTTGACAAGGGCCTATCAGTTTGATATAGCAAAACAAACA
TGGCACATTCAAAGCAATACCCCACCTCCACAACCCCCAGCTCCGCCATTAGGTTACGTGTCCGGAGCCTTGATTTCTGATTTGGAGACAG
ATGTTCCAGATGAGGATGCTGATGATGAAGAGGAACCATTAGAAATTCCCAGGCCCCTCAGAGCACTAGACCAGACACCTGGATCCAGTAT
GGACAATCTAGACAGCTCTGTCACAGGAAAAGCCTTTAGCTCCTCTCAAAGGCAGCGGCCCACCAGCCCATTTTCTACGGACAGTAACACC
AGTGCTGCCCAGAATCAAAGCCAGAGGCCTCGGCCCACGAAAAAACATAAGGGAGGACGGATGGACCCACAGCCAGTGTTACCTCATCGGA

No.15

GGGAAGGGATGCCAGATGAGGAGTCCCTGGTGCCCTACAGCAAGCCCAGCTTCCCGTCTCCTGGCGGGCACAGCTCATCAGGAACATCTTC
CTCTAAAGGGTCCACTGGCCCTCGGAAGGCTGACGTGCTGAGAGGCAGCCACCAGAGGAATGCCAATGACCTTCTCGACATTGGATATGTG
GGCTCAAATAGTCAAGGACAGTTTACAGAGTAACTGAGAGGGGACATACAAAGATGCTCTGAGGACCATCAGGTCCAAACTCATGGAAGTG
ATAACATTTAACAGTGCAATGAACAATTTCTTTATTTATGTACTATTAAAAAAAAACTGTAAATGCAATGTAAAGACACACAGCCACACAT
ATTCCACAGTTATTTTCTTGTGTTCTTCTCTTACTTACACCACCCACCTTAACTCTTTCTTGTCAGGAGTATATAAGAAAAAGAAAGAAAG
CAAAACTCACCCTCCAGGAAGAAAAGGATTTTTCCTCTGTATATAATTTCTTTTGTGCATTGCTATGCAAGCTCACTCTTTTTAACTCTGT
TCATATTATTGTCTGTTCTGATTGGTCTGTTGTACTATATGTGAATTAATGGGCTGTGGTGCCATATATTAACTTTTAATTGTGTAACTTT
TTATGTTTAAATTTTGCACTGCGATTTTATTTGGTGATAAGCACAAATCTCTACTCCTCATGACATGAAGAAAAGATTGAATGTGAAGGGA

No.16

GTTTTCTGTACTGTAAGTTAGGTTGGATAATGCTGGTGTAACCAATCCAGTTAGATGGTTTTCAGTTGGGGGTGTAGAAATAGGAAGATCG
AAGGAATGATGGTGTTGGCAAAGTCTTCTTGAAACAACAGATATTGAGACAATTTTAAGAAGCAGAAAGATGGATACTATTGACTAAAGCA
GGGGTCAAAAGAAGGGGGTTTAAGTCTAGACAGAGTATGTAATAAAGTATGGTGGTAGCAAAGATGTACTAACTTGCTTTAAAAATATTAT
TAAAGTTTTATTTAGAATGAACTTTACCTGCCATTGTAATTAACCCATCTTAGAATTACAATGAGCAAAATAAAAATCAAGGTGTTTCAAG

No.14

AGATCTGTATTTATATTACAGCTTTTTAAAAACAGCAGTTTTGAATAATCATAATTAATCTCTCCAACTCATTAAGTCAGACTATAACATG
GAGTTCCCCAAGGAAACAAACAAAAAAAATGAACTCCAGAAACTCTAGCAAATAGTTAAAGAAGCAACTTATTCTTCGAGCATACTCGGCT
ACTCTTGAATATGAGCTCAATTGCAATATCTTGTTCATCTTTCTAATACATGTACAGTACATATTAGAGGATGTAGGCACATCACTTAAGT
TCATTTCTTTTTAAAAAAAAAAAGCAAACTAATGCATTCTGCATACAATTTTGTGTTTCTTTTGGTTAAAATGTGAAATCGATGCCACCAG
ATGTATAGCCACATTATAAGTGCTTAAAGGCAGTGCTCACAGTATGTTCAGTTTCCAGTTAGTAGATTTATTGGACTCAATATTTTCTGGT
ACATTCTCAGAAATTGGCTACCAACTCAAATCTGTTTGACCCATTTTGAATGAGTAAATCGGAAAGAACAGTCAAAAGGGAGAAAAAAATG
CATGTTTATACACTTTTTAAATAAAACCAAATCTTGTAAGTGGACATTAATAACAGTGTCCTGCCTCATTTGTTTTCACGACTCTGAGAAC
AAGGAGTTGCTCAACATCCATTATGTATGCTCAAGATAAATGTGACTTTGAAATGTATGTTAGCTACTGTACATGTATGTAAGTCAAGTAG
ATGATAGCCTTCCTGAAACTACCACTAATAATATTTTCCCATGTGCTACTTACACAAGCTTAAATTCTAGTTCAGCAAATGTTTTTCATTA
TTTTTTCTTACGGTTTATTCAATAGAGGGTACCTAGAAAAATATTCAGTGTGATGGTCAACATATTGAGTGTTCAAGTAGGGGATATTTAC
TTTGGCAATACATACTGTGCATTATTGGTTCAGATGCCAATTCATGAGACATTATCATAGGAAATGGATAATATAAGAACATGGCTCTAAG
GGTCCTTCTCCACAGGCAAGCATAGAATTTGAGGTTCTTCATGGGTCCCAAAACTTGCTTAGTGTTATATACTTAACTTCCTTCACTGAAG
TATCAAGCAATTCTAACACACCGTTTAACAACTTTCTACCATCAGTGCCTGCATTTTATTGAAGTTTGATGTCTCTGGTACAAGGAAATAG
AGAAAGCTGAACAGTTTCTGTTGGGAGAAAAAGTTTCTGTAATTATTGTCAATTCTAAGTTGTGATGAATATAACTTTTCATCTATACAGT
TCCTTACTCATGGTCTCTATTTAAAAGCATAGGATCTTTTAAGATATACTTTTAGCTGAGTCTTCACTGTATGTTCATGAATAATATAATT
TAGTGAAAAACTTTAGTACATCATATGTATGGTAACTGACATTTGCAGATTATTATCAAACCAAATCATAAGAAATAATATCTTCAGATCA
CATGATGGCTTACCGAACAATTTCAGGTTATAGCATGATGTCTTTCTGATTGTCTTCAGTATATGATCATAAGCAATTATATTGTGTCCAT
ACAGTTTACATACTGTGATTCCCGCAAAATCTTGCATCTGATATCTACCATGAAGATGAAATAATATGCTATGAATTTCAGGCATAGACAT
ACTATCAGCAATAGAAAGCCTTCATCATCTAGATATCTGGATAAATCCATTCATCGTCAACCATCAGATTGTTAACTGGGGAACACGTACC
TCCCTCACAAATGCTTTCTTTTCGTCTTGTAGTTGATTGAATGTGATAGGAAATCTGCTGAGTTGACTCTTCTGTGGGCCTGAGCAGAAAC

ATGCATAAGCCATCAGAAGGAAGTCTGGATGGGAAGGAAACCTTGTCTTTCACTCGTAGGCAATACTGCTGTGTAGTCACGGCAGACCCAA
GCACAGCTTCTTTCAGAACATGCAAAGTCTCCTTATGAAATATACACATCAGTGCCCAAACTCAGGTGTACCAGTGTAAAGCGTAGGGTTT
GCCTACCCTTTTTTCATATCATTTTACATAAGCAAGTTAGGAGACTACAGGGCTATGAAAATAAATGTTTCCTTAACAGGATTTGGAGCAA
GTTCACGGGAGTACATAGGACAGTCGAGCCCAGTGTAAACACTACCAAAGGAAAGGACAGCCTTGCAGAAGCGTGAGCTCCGACTAACAGA
GCGGCTTAAGTATGCTCTGTGATACAGCTTAATGCACTCCCTATTCAAATCCCACTGTCAATGGGATTTAAGTTGCCAGTTTATATGTGAT
TTCATGGAACACAACCAAGTTCTACTTAATACCTTTATGCAATTTTTTTTTCACTGATGAAAAATCAATATCCACATCAATGGGTGCCCAG
CTGCGGATCACCTCTCCTAAAAAGCGATCTTGTCAGTATTGAAGAAATGTAGAACCATAATTTTCACTACTAATATCATTTTGCTCAACTA
CTTCTGTATATTGTGGAAAAAGGGGTGGGGCCGGGGGAATGTGTGTTTCCACATTTTCTGAGAGGTGAAAATATTTTTGTGTTGCTTTGTC
CTAATGGGTACTTCAAATGAAACACCATACAATTTTAAATGACATGCAATTAACGAATATTTTTTACTCTCATAACATTAAAAGCATTTGT
TGTG

Positions of primers are marked in gray.

## Supplementary figure 11. Mouse *Robo2b* variant 3.

No.3

ATGAATCCTCTGATGTTTACACTATTATTGCTCTTTGGATTTCTCTGCATTCAGATTGATGGATCTCGTCTTCGTCAAGAAGACTTTCCCC
CCAGGATCGTAGAACACCCTTCTGATGTCATCGTCTCCAAGGGAGAGCCCACCACTCTGAACTGTAAAGCAGAGGGCCGACCCACCCCCAC
CATTGAATGGTACAAGGATGGTGAGAGGGTGGAGACAGACAAGGATGATCCCAGGTCCCACAGAATGCTTCTGCCCAGCGGATCTTTATTC
TTTTTGCGAATTGTTCATGGGCGCAGAAGTAAACCGGACGAAGGGAGTTACGTTTGTGTTGCAAGGAACTATCTTGGTGAAGCAGTGAGTC

No.5

GAAATGCATCTCTGGAAGTGGCATTATTGCGAGATGACTTCCGGCAAAACCCCACAGATGTGGTAGTCGCAGCTGGAGAGCCTGCAATCTT
GGAGTGCCAGCCACCACGGGGACACCCAGAACCAACCATCTACTGGAAAAAGGACAAAGTCCGAATTGATGACAAGGAAGAGAGAATAAGT
ATCCGTGGTGGGAAGCTGATGATCTCTAATACTAGGAAAAGCGATGCTGGCATGTACACCTGTGTGGGAACCAATATGGTGGGAGAAAGGG
ACAGCGACCCTGCAGAGCTCACTGTCTTTGAACGACCCACATTTCTCAGGAGGCCAATTAACCAGGTGGTGCTAGAGGAAGAAGCTGTAGA
ATTCCGTTGTCAGGTCCAAGGAGATCCCCAGCCAACGGTGAGGTGGAAAAAAGATGATGCAGACTTGCCGAGAGGAAGGTATGATATCAAA
GATGACTACACGCTGAGAATTAAAAAGGCCATGAGTACTGATGAAGGTACCTATGTGTGTATTGCTGAGAATCGGGTGGGAAAAGTGGAAG

No.6

CCTCTGCTACCCTCACTGTCCGAGCTCCTCCACAGTTTGTGGTTAGGCCAAGAGATCAGATCGTTGCTCAAGGCCGAACAGTGACATTCCC
CTGTGAAACTAAAGGAAACCCACAGCCAGCTGTTTTTTGGCAGAAAGAAGGCAGCCAGAACCTACTTTTCCCGAATCAACCTCAGCAGCCC
AACAGCCGATGTTCAGTGTCGCCCACGGGGGACCTCACCATCACCAACATCCAGCGTTCAGATGCGGGTTACTACATCTGCCAGGCCCTAA
CCGTGGCAGGAAGCATTTTAGCTAAAGCACAGTTGGAAGTTACTGACGTTTTGACAGATAGACCTCCACCCATAATCTTGCAAGGACCAAT
AAACCAAACACTTGCAGTAGACGGTACAGCATTGTTGAAGTGTAAAGCCACTGGTGAGCCTCTGCCTGTAATTAGCTGGCTAAAGGAGGGC

No.1(rev)+ No.7(for)

TTTACTTTTCTGGGGAGAGATCCAAGAGCCACGATCCAAGACCAAGGAACACTGCAGATTAAGAATTTACGGATATCTGATACTGGCACTT
ATACTTGTGTGGCTACAAGTTCCAGTGGAGAGACTTCCTGGAGTGCAGTGCTGGATGTAACAGAATCTGGAGCAACAATCAGTAAAAATTA
TGATATGAATGACCTCCCGGGACCACCATCCAAACCTCAGGTCACTGATGTTTCTAAGAACAGTGTCACCTTATCCTGGCAGCCAGGTACA

No.9

CCTGGCGTTCTTCCTGCAAGCGCGTATATCATTGAGGCTTTCAGCCAATCGGTGAGCAATAGCTGGCAGACAGTGGCAAACCATGTTAAGA
CAACTCTGTATACAGTAAGGGGGCTGAGGCCCAACACAATCTACTTGTTTATGGTCAGAGCGATCAACCCACAAGGTCTCAGTGATCCAAG
TCCTATGTCGGATCCTGTACGCACACAAGATATCAGCCCCCCAGCACAAGGAGTGGACCACAGACAGGTGCAGAAGGAATTAGGTGATGTC
GTTGTTCGTCTCCATAATCCAGTTGTCCTGACACCTACAACTGTTCAAGTCACATGGACGGTGGACCGACAACCCCAGTTTATTCAGGGCT
ACAGAGTGATGTACCGTCAGACTTCGGGACTACAAGCCTCAACTGTGTGGCAGAATCTAGACGCCAAAGTCCCGACTGAGAGGAGTGCTGT
CCTTGTGAATTTGAAAAAGGGGGTGACTTATGAAATTAAAGTCCGGCCGTATTTTAACGAGTTCCAAGGAATGGACAGTGAATCGAAAACA
GTCCGAACCACTGAGGAAGCCCCAAGTGCCCCTCCCCAGTCTGTCACTGTGCTGACAGTTGGAAGTCACAACAGCACAAGCATCAGTGTTT
CCTGGGATCCTCCACCAGCCGACCACCAGAATGGAATTATTCAGGAATATAAGATCTGGTGTCTGGGAAACGAAACGCGATTCCATATCAA

No.10

TAAAACGGTGGATGCAGCCATTCGCTCTGTAGTAATAGGTGGCTTGTTCCCTGGAATTCAGTACCGGGTAGAAGTGGCAGCTAGCACAAGT
GCAGGGGTTGGAGTAAAAAGTGAACCACAGCCGATAATAATTGGGGGACGTAACGAAGTTGTCATTACTGAAAACAATAACAGCATCACTG

No.11(for)+No.8(rev)

AGCAAATCACGGATGTCGTGAAGCAACCGGCATTTATAGCTGGCATTGGTGGTGCCTGCTGGGTAATTCTGATGGGTTTTAGCATCTGGTT
GTACTGGAGAAGAAAGAAGAGAAAGGGACTCAGCAATTATGCTGTAACATTTCAAAGAGGAGATGGAGGACTAATGAGCAATGGGAGCCGT
CCAGGTCTTCTAAATGCTGGCGATCCCAATTACCCATGGCTTGCTGATTCTTGGCCAGCCACGAGTTTGCCAGTGAACAATAGCAATAGTG
GCCCAAATGAAATTGGAAATTTTGGGCGTGGAGATGTGCTGCCTCCGGTGCCAGGCCAAGGGGATAAAACAGCGACCATGCTCTCGGATGG
AGCCATTTATAGCAGCATTGACTTCACTACCAAAACCACTTACAACAGTTCCAGCCAAATAACACAGGCCACCCCATATGCCACTACACAA
ATCCTGCATTCAAACAGCATCCACGAACTGGCAGTTGATCTTCCTGATCCACAGTGGAAAAGCTCAGTTCAACAGAAGACAGACCTCATGG
GATTTGGTTATTCGCTACCTGATCAGAACAAGGGGAACAACGCCtTACTttACATCCcTGACTACcGATtGGCTGAGGGATTGTCTAATAG

AATGCCACACAACCAGTCACAGGATTTCAGCACCACCAGCTCTCACAACAGCTCAGAAAGGAGTGGCAGTCTCTCAGGTGGGAAAGGTGGA
AAAAAGAAGAAAACTAAAAATTCTTCGAAAGCGCAGAAAAACAACGGATCCACTTGGGCTAATGTCCCTCTACCTCCTCCTCCCGTCCAGC
CCCTTCCTGGTACAGAGCTGGGCCACTATGCTGCGGAACAAGAAAATGGCTATGACAGTGACAGCTGGTGCCCACCATTACCGGTGCAAAC
ATACTTGCATCAGGGTATGGAAGATGAGCTGGAAGAAGACGAAGATCGGGTCCCAACACCTCCTGTTCGCGGCGTGGCCTCTTCACCAGCT

No.13(for)+ No.12(rev)

ATCTCTTTTGGACAGCAGTCCACTGCCACTCTTACTCCATCCCCACGGGAAGAGATGCAACCCATGCTGCAAGCTCACTTGGATGAGTTGA
CAAGGGCCTATCAGTTTGATATAGCAAAACAAACATGGCACATTCAAAGCAATACCCCACCTCCACAACCCCCAGCTCCGCCATTAGGTTA
CGTGTCCGGAGCCTTGATTTCTGATTTGGAGACAGATGTTCCAGATGAGGATGCTGATGATGAAGAGGAACCATTAGAAATTCCCAGGCCC
CTCAGAGCACTAGACCAGACACCTGGATCCAGTATGGACAATCTAGACAGCTCTGTCACAGGAAAAGCCTTTAGCTCCTCTCAAAGGCAGC
GGCCCACCAGCCCATTTTCTACGGACAGTAACACCAGTGCTGCCCAGAATCAAAGCCAGAGGCCTCGGCCCACGAAAAAACATAAGGGAGG
ACGGATGGACCCACAGCCAGTGTTACCTCATCGGAGGGAAGGGATGCCAGATGATCTTCCCCCACCACCAGATCCCCCACCAGGTCAGGGT
TTAAGGCAGCAGATAGGCCTGAGCCAACACTCTGGTAACGTGGAAAACTCAACAGAGAGAAAAGGAAGCTCTCTAGAGAGACAACAAGCAG
CCAACTTAGAAGACACAAAGAGCTCATTGGATTGTCCAGCTAAAACTGTTCTAGAGTGGCAGCGTCAAACCCAGGACTGGATAAACTCTAC

No.15

AGAACGCCAAGAAGAAACTCGGAAAGCCCCACACAAACAAGGGGTTGGATCAGAGGAGTCCCTGGTGCCCTACAGCAAGCCCAGCTTCCCG
TCTCCTGGCGGGCACAGCTCATCAGGAACATCTTCCTCTAAAGGGTCCACTGGCCCTCGGAAGGCTGACGTGCTGAGAGGCAGCCACCAGA
GGAATGCCAATGACCTTCTCGACATTGGATATGTGGGCTCAAATAGTCAAGGACAGTTTACAGAGTAACTGAGAGGGGACATACAAAGATG
CTCTGAGGACCATCAGGTCCAAACTCATGGAAGTGATAACATTTAACAGTGCAATGAACAATTTCTTTATTTATGTACTATTAAAAAAAAA
CTGTAAATGCAATGTAAAGACACACAGCCACACATATTCCACAGTTATTTTCTTGTGTTCTTCTCTTACTTACACCACCCACCTTAACTCT
TTCTTGTCAGGAGTATATAAGAAAAAGAAAGAAAGCAAAACTCACCCTCCAGGAAGAAAAGGATTTTTCCTCTGTATATAATTTCTTTTGT
GCATTGCTATGCAAGCTCACTCTTTTTAACTCTGTTCATATTATTGTCTGTTCTGATTGGTCTGTTGTACTATATGTGAATTAATGGGCTG
TGGTGCCATATATTAACTTTTAATTGTGTAACTTTTTATGTTTAAATTTTGCACTGCGATTTTATTTGGTGATAAGCACAAATCTCTACTC

No.16

CTCATGACATGAAGAAAAGATTGAATGTGAAGGGAGTTTTCTGTACTGTAAGTTAGGTTGGATAATGCTGGTGTAACCAATCCAGTTAGAT
GGTTTTCAGTTGGGGGTGTAGAAATAGGAAGATCGAAGGAATGATGGTGTTGGCAAAGTCTTCTTGAAACAACAGATATTGAGACAATTTT
AAGAAGCAGAAAGATGGATACTATTGACTAAAGCAGGGGTCAAAAGAAGGGGGTTTAAGTCTAGACAGAGTATGTAATAAAGTATGGTGGT
AGCAAAGATGTACTAACTTGCTTTAAAAATATTATTAAAGTTTTATTTAGAATGAACTTTACCTGCCATTGTAATTAACCCATCTTAGAAT
TACAATGAGCAAAATAAAAATCAAGGTGTTTCAAGAGATCTGTATTTATATTACAGCTTTTTAAAAACAGCAGTTTTGAATAATCATAATT

No.14

AATCTCTCCAACTCATTAAGTCAGACTATAACATGGAGTTCCCCAAGGAAACAAACAAAAAAAATGAACTCCAGAAACTCTAGCAAATAGT
TAAAGAAGCAACTTATTCTTCGAGCATACTCGGCTACTCTTGAATATGAGCTCAATTGCAATATCTTGTTCATCTTTCTAATACATGTACA
GTACATATTAGAGGATGTAGGCACATCACTTAAGTTCATTTCTTTTTAAAAAAAAAAAGCAAACTAATGCATTCTGCATACAATTTTGTGT
TTCTTTTGGTTAAAATGTGAAATCGATGCCACCAGATGTATAGCCACATTATAAGTGCTTAAAGGCAGTGCTCACAGTATGTTCAGTTTCC
AGTTAGTAGATTTATTGGACTCAATATTTTCTGGTACATTCTCAGAAATTGGCTACCAACTCAAATCTGTTTGACCCATTTTGAATGAGTA
AATCGGAAAGAACAGTCAAAAGGGAGAAAAAAATGCATGTTTATACACTTTTTAAATAAAACCAAATCTTGTAAGTGGACATTAATAACAG
TGTCCTGCCTCATTTGTTTTCACGACTCTGAGAACAAGGAGTTGCTCAACATCCATTATGTATGCTCAAGATAAATGTGACTTTGAAATGT
ATGTTAGCTACTGTACATGTATGTAAGTCAAGTAGATGATAGCCTTCCTGAAACTACCACTAATAATATTTTCCCATGTGCTACTTACACA
AGCTTAAATTCTAGTTCAGCAAATGTTTTTCATTATTTTTTCTTACGGTTTATTCAATAGAGGGTACCTAGAAAAATATTCAGTGTGATGG
TCAACATATTGAGTGTTCAAGTAGGGGATATTTACTTTGGCAATACATACTGTGCATTATTGGTTCAGATGCCAATTCATGAGACATTATC
ATAGGAAATGGATAATATAAGAACATGGCTCTAAGGGTCCTTCTCCACAGGCAAGCATAGAATTTGAGGTTCTTCATGGGTCCCAAAACTT
GCTTAGTGTTATATACTTAACTTCCTTCACTGAAGTATCAAGCAATTCTAACACACCGTTTAACAACTTTCTACCATCAGTGCCTGCATTT
TATTGAAGTTTGATGTCTCTGGTACAAGGAAATAGAGAAAGCTGAACAGTTTCTGTTGGGAGAAAAAGTTTCTGTAATTATTGTCAATTCT
AAGTTGTGATGAATATAACTTTTCATCTATACAGTTCCTTACTCATGGTCTCTATTTAAAAGCATAGGATCTTTTAAGATATACTTTTAGC
TGAGTCTTCACTGTATGTTCATGAATAATATAATTTAGTGAAAAACTTTAGTACATCATATGTATGGTAACTGACATTTGCAGATTATTAT

CAAACCAAATCATAAGAAATAATATCTTCAGATCACATGATGGCTTACCGAACAATTTCAGGTTATAGCATGATGTCTTTCTGATTGTCTT
CAGTATATGATCATAAGCAATTATATTGTGTCCATACAGTTTACATACTGTGATTCCCGCAAAATCTTGCATCTGATATCTACCATGAAGA
TGAAATAATATGCTATGAATTTCAGGCATAGACATACTATCAGCAATAGAAAGCCTTCATCATCTAGATATCTGGATAAATCCATTCATCG
TCAACCATCAGATTGTTAACTGGGGAACACGTACCTCCCTCACAAATGCTTTCTTTTCGTCTTGTAGTTGATTGAATGTGATAGGAAATCT
GCTGAGTTGACTCTTCTGTGGGCCTGAGCAGAAACATGCATAAGCCATCAGAAGGAAGTCTGGATGGGAAGGAAACCTTGTCTTTCACTCG
TAGGCAATACTGCTGTGTAGTCACGGCAGACCCAAGCACAGCTTCTTTCAGAACATGCAAAGTCTCCTTATGAAATATACACATCAGTGCC
CAAACTCAGGTGTACCAGTGTAAAGCGTAGGGTTTGCCTACCCTTTTTTCATATCATTTTACATAAGCAAGTTAGGAGACTACAGGGCTAT
GAAAATAAATGTTTCCTTAACAGGATTTGGAGCAAGTTCACGGGAGTACATAGGACAGTCGAGCCCAGTGTAAACACTACCAAAGGAAAGG
ACAGCCTTGCAGAAGCGTGAGCTCCGACTAACAGAGCGGCTTAAGTATGCTCTGTGATACAGCTTAATGCACTCCCTATTCAAATCCCACT
GTCAATGGGATTTAAGTTGCCAGTTTATATGTGATTTCATGGAACACAACCAAGTTCTACTTAATACCTTTATGCAATTTTTTTTTCACTG
ATGAAAAATCAATATCCACATCAATGGGTGCCCAGCTGCGGATCACCTCTCCTAAAAAGCGATCTTGTCAGTATTGAAGAAATGTAGAACC
ATAATTTTCACTACTAATATCATTTTGCTCAACTACTTCTGTATATTGTGGAAAAAGGGGTGGGGCCGGGGGAATGTGTGTTTCCACATTT
TCTGAGAGGTGAAAATATTTTTGTGTTGCTTTGTCCTAATGGGTACTTCAAATGAAACACCATACAATTTTAAATGACATGCAATTAACGA
ATATTTTTTACTCTCATAACATTAAAAGCATTTGTTGTG

Positions of primers are marked in gray.

## Supplementary figure 12. Mouse *Robo2b* variant 4.

No.3

ATGAATCCTCTGATGTTTACACTATTATTGCTCTTTGGATTTCTCTGCATTCAGATTGATGGATCTCGTCTTCGTCAAGAAGACTTTCCCC
CCAGGATCGTAGAACACCCTTCTGATGTCATCGTCTCCAAGGGAGAGCCCACCACTCTGAACTGTAAAGCAGAGGGCCGACCCACCCCCAC
CATTGAATGGTACAAGGATGGTGAGAGGGTGGAGACAGACAAGGATGATCCCAGGTCCCACAGAATGCTTCTGCCCAGCGGATCTTTATTC
TTTTTGCGAATTGTTCATGGGCGCAGAAGTAAACCGGACGAAGGGAGTTACGTTTGTGTTGCAAGGAACTATCTTGGTGAAGCAGTGAGTC

No.5

GAAATGCATCTCTGGAAGTGGCATTATTGCGAGATGACTTCCGGCAAAACCCCACAGATGTGGTAGTCGCAGCTGGAGAGCCTGCAATCTT
GGAGTGCCAGCCACCACGGGGACACCCAGAACCAACCATCTACTGGAAAAAGGACAAAGTCCGAATTGATGACAAGGAAGAGAGAATAAGT
ATCCGTGGTGGGAAGCTGATGATCTCTAATACTAGGAAAAGCGATGCTGGCATGTACACCTGTGTGGGAACCAATATGGTGGGAGAAAGGG
ACAGCGACCCTGCAGAGCTCACTGTCTTTGAACGACCCACATTTCTCAGGAGGCCAATTAACCAGGTGGTGCTAGAGGAAGAAGCTGTAGA
ATTCCGTTGTCAGGTCCAAGGAGATCCCCAGCCAACGGTGAGGTGGAAAAAAGATGATGCAGACTTGCCGAGAGGAAGGTATGATATCAAA
GATGACTACACGCTGAGAATTAAAAAGGCCATGAGTACTGATGAAGGTACCTATGTGTGTATTGCTGAGAATCGGGTGGGAAAAGTGGAAG

No.6

CCTCTGCTACCCTCACTGTCCGAGCTCCTCCACAGTTTGTGGTTAGGCCAAGAGATCAGATCGTTGCTCAAGGCCGAACAGTGACATTCCC
CTGTGAAACTAAAGGAAACCCACAGCCAGCTGTTTTTTGGCAGAAAGAAGGCAGCCAGAACCTACTTTTCCCGAATCAACCTCAGCAGCCC
AACAGCCGATGTTCAGTGTCGCCCACGGGGGACCTCACCATCACCAACATCCAGCGTTCAGATGCGGGTTACTACATCTGCCAGGCCCTAA
CCGTGGCAGGAAGCATTTTAGCTAAAGCACAGTTGGAAGTTACTGACGTTTTGACAGATAGACCTCCACCCATAATCTTGCAAGGACCAAT
AAACCAAACACTTGCAGTAGACGGTACAGCATTGTTGAAGTGTAAAGCCACTGGTGAGCCTCTGCCTGTAATTAGCTGGCTAAAGGAGGGC

No.1(rev)+ No.7(for)

TTTACTTTTCTGGGGAGAGATCCAAGAGCCACGATCCAAGACCAAGGAACACTGCAGATTAAGAATTTACGGATATCTGATACTGGCACTT
ATACTTGTGTGGCTACAAGTTCCAGTGGAGAGACTTCCTGGAGTGCAGTGCTGGATGTAACAGAATCTGGAGCAACAATCAGTAAAAATTA
TGATATGAATGACCTCCCGGGACCACCATCCAAACCTCAGGTCACTGATGTTTCTAAGAACAGTGTCACCTTATCCTGGCAGCCAGGTACA

No.9

CCTGGCGTTCTTCCTGCAAGCGCGTATATCATTGAGGCTTTCAGCCAATCGGTGAGCAATAGCTGGCAGACAGTGGCAAACCATGTTAAGA
CAACTCTGTATACAGTAAGGGGGCTGAGGCCCAACACAATCTACTTGTTTATGGTCAGAGCGATCAACCCACAAGGTCTCAGTGATCCAAG
TCCTATGTCGGATCCTGTACGCACACAAGATATCAGCCCCCCAGCACAAGGAGTGGACCACAGACAGGTGCAGAAGGAATTAGGTGATGTC
GTTGTTCGTCTCCATAATCCAGTTGTCCTGACACCTACAACTGTTCAAGTCACATGGACGGTGGACCGACAACCCCAGTTTATTCAGGGCT
ACAGAGTGATGTACCGTCAGACTTCGGGACTACAAGCCTCAACTGTGTGGCAGAATCTAGACGCCAAAGTCCCGACTGAGAGGAGTGCTGT
CCTTGTGAATTTGAAAAAGGGGGTGACTTATGAAATTAAAGTCCGGCCGTATTTTAACGAGTTCCAAGGAATGGACAGTGAATCGAAAACA
GTCCGAACCACTGAGGAAGCCCCAAGTGCCCCTCCCCAGTCTGTCACTGTGCTGACAGTTGGAAGTCACAACAGCACAAGCATCAGTGTTT
CCTGGGATCCTCCACCAGCCGACCACCAGAATGGAATTATTCAGGAATATAAGATCTGGTGTCTGGGAAACGAAACGCGATTCCATATCAA

No.10

TAAAACGGTGGATGCAGCCATTCGCTCTGTAGTAATAGGTGGCTTGTTCCCTGGAATTCAGTACCGGGTAGAAGTGGCAGCTAGCACAAGT
GCAGGGGTTGGAGTAAAAAGTGAACCACAGCCGATAATAATTGGGGGACGTAACGAAGTTGTCATTACTGAAAACAATAACAGCATCACTG

No.11(for)+No.8(rev)

AGCAAATCACGGATGTCGTGAAGCAACCGGCATTTATAGCTGGCATTGGTGGTGCCTGCTGGGTAATTCTGATGGGTTTTAGCATCTGGTT
GTACTGGAGAAGAAAGAAGAGAAAGGGACTCAGCAATTATGCTGTAACATTTCAAAGAGGAGATGGAGGACTAATGAGCAATGGGAGCCGT
CCAGGTCTTCTAAATGCTGGCGATCCCAATTACCCATGGCTTGCTGATTCTTGGCCAGCCACGAGTTTGCCAGTGAACAATAGCAATAGTG
GCCCAAATGAAATTGGAAATTTTGGGCGTGGAGATGTGCTGCCTCCGGTGCCAGGCCAAGGGGATAAAACAGCGACCATGCTCTCGGATGG
AGCCATTTATAGCAGCATTGACTTCACTACCAAAACCACTTACAACAGTTCCAGCCAAATAACACAGGCCACCCCATATGCCACTACACAA
ATCCTGCATTCAAACAGCATCCACGAACTGGCAGTTGATCTTCCTGATCCACAGTGGAAAAGCTCAGTTCAACAGAAGACAGACCTCATGG
GATTTGGTTATTCGCTACCTGATCAGAACAAGGGGAACAACGCCtTACTttACATCCcTGACTACcGATtGGCTGAGGGATTGTCTAATAG

AATGCCACACAACCAGTCACAGGATTTCAGCACCACCAGCTCTCACAACAGCTCAGAAAGGAGTGGCAGTCTCTCAGGTGGGAAAGGTGGA
AAAAAGAAGAAAACTAAAAATTCTTCGAAAGCGCAGAAAAACAACGGATCCACTTGGGCTAATGTCCCTCTACCTCCTCCTCCCGTCCAGC
CCCTTCCTGGTACAGAGCTGGGCCACTATGCTGCGGAACAAGAAAATGGCTATGACAGTGACAGCTGGTGCCCACCATTACCGGTGCAAAC
ATACTTGCATCAGGGTATGGAAGATGAGCTGGAAGAAGACGAAGATCGGGTCCCAACACCTCCTGTTCGCGGCGTGGCCTCTTCACCAGCT

No.13(for)+ No.12(rev)

ATCTCTTTTGGACAGCAGTCCACTGCCACTCTTACTCCATCCCCACGGGAAGAGATGCAACCCATGCTGCAAGCTCACTTGGATGAGTTGA
CAAGGGCCTATCAGTTTGATATAGCAAAACAAACATGGCACATTCAAAGCAATACCCCACCTCCACAACCCCCAGCTCCGCCATTAGGTTA
CGTGTCCGGAGCCTTGATTTCTGATTTGGAGACAGATGTTCCAGATGAGGATGCTGATGATGAAGAGGAACCATTAGAAATTCCCAGGCCC
CTCAGAGCACTAGACCAGACACCTGGATCCAGTATGGACAATCTAGACAGCTCTGTCACAGGAAAAGCCTTTAGCTCCTCTCAAAGGCAGC
GGCCCACCAGCCCATTTTCTACGGACAGTAACACCAGTGCTGCCCAGAATCAAAGCCAGAGGCCTCGGCCCACGAAAAAACATAAGGGAGG

No.15

ACGGATGGACCCACAGCCAGTGTTACCTCATCGGAGGGAAGGGATGCCAGATGAGGAGTCCCTGGTGCCCTACAGCAAGCCCAGCTTCCCG
TCTCCTGGCGGGCACAGCTCATCAGGAACATCTTCCTCTAAAGGGTCCACTGGCCCTCGGAAGGCTGACGTGCTGAGAGGCAGCCACCAGA
GGAATGCCAATGACCTTCTCGACATTGGATATGTGGGCTCAAATAGTCAAGGACAGTTTACAGAGTAACTGAGAGGGGACATACAAAGATG
CTCTGAGGACCATCAGGTCCAAACTCATGGAAGTGATAACATTTAACAGTGCAATGAACAATTTCTTTATTTATGTACTATTAAAAAAAAA
CTGTAAATGCAATGTAAAGACACACAGCCACACATATTCCACAGTTATTTTCTTGTGTTCTTCTCTTACTTACACCACCCACCTTAACTCT
TTCTTGTCAGGAGTATATAAGAAAAAGAAAGAAAGCAAAACTCACCCTCCAGGAAGAAAAGGATTTTTCCTCTGTATATAATTTCTTTTGT
GCATTGCTATGCAAGCTCACTCTTTTTAACTCTGTTCATATTATTGTCTGTTCTGATTGGTCTGTTGTACTATATGTGAATTAATGGGCTG
TGGTGCCATATATTAACTTTTAATTGTGTAACTTTTTATGTTTAAATTTTGCACTGCGATTTTATTTGGTGATAAGCACAAATCTCTACTC

No.16

CTCATGACATGAAGAAAAGATTGAATGTGAAGGGAGTTTTCTGTACTGTAAGTTAGGTTGGATAATGCTGGTGTAACCAATCCAGTTAGAT
GGTTTTCAGTTGGGGGTGTAGAAATAGGAAGATCGAAGGAATGATGGTGTTGGCAAAGTCTTCTTGAAACAACAGATATTGAGACAATTTT
AAGAAGCAGAAAGATGGATACTATTGACTAAAGCAGGGGTCAAAAGAAGGGGGTTTAAGTCTAGACAGAGTATGTAATAAAGTATGGTGGT
AGCAAAGATGTACTAACTTGCTTTAAAAATATTATTAAAGTTTTATTTAGAATGAACTTTACCTGCCATTGTAATTAACCCATCTTAGAAT
TACAATGAGCAAAATAAAAATCAAGGTGTTTCAAGAGATCTGTATTTATATTACAGCTTTTTAAAAACAGCAGTTTTGAATAATCATAATT

No.14

AATCTCTCCAACTCATTAAGTCAGACTATAACATGGAGTTCCCCAAGGAAACAAACAAAAAAAATGAACTCCAGAAACTCTAGCAAATAGT
TAAAGAAGCAACTTATTCTTCGAGCATACTCGGCTACTCTTGAATATGAGCTCAATTGCAATATCTTGTTCATCTTTCTAATACATGTACA
GTACATATTAGAGGATGTAGGCACATCACTTAAGTTCATTTCTTTTTAAAAAAAAAAAGCAAACTAATGCATTCTGCATACAATTTTGTGT
TTCTTTTGGTTAAAATGTGAAATCGATGCCACCAGATGTATAGCCACATTATAAGTGCTTAAAGGCAGTGCTCACAGTATGTTCAGTTTCC
AGTTAGTAGATTTATTGGACTCAATATTTTCTGGTACATTCTCAGAAATTGGCTACCAACTCAAATCTGTTTGACCCATTTTGAATGAGTA
AATCGGAAAGAACAGTCAAAAGGGAGAAAAAAAATGCATGTTTATACACTTTTTAAATAAAACCAAATCTTGTAAGTGGACATTAATAACAG
TGTCCTGCCTCATTTGTTTTCACGACTCTGAGAACAAGGAGTTGCTCAACATCCATTATGTATGCTCAAGATAAATGTGACTTTGAAATGT
ATGTTAGCTACTGTACATGTATGTAAGTCAAGTAGATGATAGCCTTCCTGAAACTACCACTAATAATATTTTCCCATGTGCTACTTACACA
AGCTTAAATTCTAGTTCAGCAAATGTTTTTCATTATTTTTTCTTACGGTTTATTCAATAGAGGGTACCTAGAAAAATATTCAGTGTGATGG
TCAACATATTGAGTGTTCAAGTAGGGGATATTTACTTTGGCAATACATACTGTGCATTATTGGTTCAGATGCCAATTCATGAGACATTATC
ATAGGAAATGGATAATATAAGAACATGGCTCTAAGGGTCCTTCTCCACAGGCAAGCATAGAATTTGAGGTTCTTCATGGGTCCCAAAACTT
GCTTAGTGTTATATACTTAACTTCCTTCACTGAAGTATCAAGCAATTCTAACACACCGTTTAACAACTTTCTACCATCAGTGCCTGCATTT
TATTGAAGTTTGATGTCTCTGGTACAAGGAAATAGAGAAAGCTGAACAGTTTCTGTTGGGAGAAAAAGTTTCTGTAATTATTGTCAATTCT
AAGTTGTGATGAATATAACTTTTCATCTATACAGTTCCTTACTCATGGTCTCTATTTAAAAGCATAGGATCTTTTAAGATATACTTTTAGC
TGAGTCTTCACTGTATGTTCATGAATAATATAATTTAGTGAAAAACTTTAGTACATCATATGTATGGTAACTGACATTTGCAGATTATTAT
CAAACCAAATCATAAGAAATAATATCTTCAGATCACATGATGGCTTACCGAACAATTTCAGGTTATAGCATGATGTCTTTCTGATTGTCTT
CAGTATATGATCATAAGCAATTATATTGTGTCCATACAGTTTACATACTGTGATTCCCGCAAAATCTTGCATCTGATATCTACCATGAAGA
TGAAATAATATGCTATGAATTTCAGGCATAGACATACTATCAGCAATAGAAAGCCTTCATCATCTAGATATCTGGATAAATCCATTCATCG

TCAACCATCAGATTGTTAACTGGGGAACACGTACCTCCCTCACAAATGCTTTCTTTTCGTCTTGTAGTTGATTGAATGTGATAGGAAATCT
GCTGAGTTGACTCTTCTGTGGGCCTGAGCAGAAACATGCATAAGCCATCAGAAGGAAGTCTGGATGGGAAGGAAACCTTGTCTTTCACTCG
TAGGCAATACTGCTGTGTAGTCACGGCAGACCCAAGCACAGCTTCTTTCAGAACATGCAAAGTCTCCTTATGAAATATACACATCAGTGCC
CAAACTCAGGTGTACCAGTGTAAAGCGTAGGGTTTGCCTACCCTTTTTTCATATCATTTTACATAAGCAAGTTAGGAGACTACAGGGCTAT
GAAAATAAATGTTTCCTTAACAGGATTTGGAGCAAGTTCACGGGAGTACATAGGACAGTCGAGCCCAGTGTAAACACTACCAAAGGAAAGG
ACAGCCTTGCAGAAGCGTGAGCTCCGACTAACAGAGCGGCTTAAGTATGCTCTGTGATACAGCTTAATGCACTCCCTATTCAAATCCCACT
GTCAATGGGATTTAAGTTGCCAGTTTATATGTGATTTCATGGAACACAACCAAGTTCTACTTAATACCTTTATGCAATTTTTTTTTCACTG
ATGAAAAATCAATATCCACATCAATGGGTGCCCAGCTGCGGATCACCTCTCCTAAAAAGCGATCTTGTCAGTATTGAAGAAATGTAGAACC
ATAATTTTCACTACTAATATCATTTTGCTCAACTACTTCTGTATATTGTGGAAAAAGGGGTGGGGCCGGGGGAATGTGTGTTTCCACATTT
TCTGAGAGGTGAAAATATTTTTGTGTTGCTTTGTCCTAATGGGTACTTCAAATGAAACACCATACAATTTTAAATGACATGCAATTAACGA
ATATTTTTTACTCTCATAACATTAAAAGCATTTGTTGTG

Positions of primers are marked in gray.