Supplementary Information for:

Convergent gene losses illuminate metabolic and physiological changes in herbivores and carnivores

Nikolai Hecker[1,2,3], Virag Sharma[1,2,3#] and Michael Hiller[1,2,3*]

[1]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
[2]Max Planck Institute for the Physics of Complex Systems, Dresden, Germany
[3]Center for Systems Biology Dresden, Germany

* hiller@mpi-cbg.de

[#] current affiliations: CRTD-DFG Center for Regenerative Therapies Dresden, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden; Paul Langerhans Institute Dresden (PLID) of the Helmholtz Center Munich at University Hospital Carl Gustav Carus and Faculty of Medicine, Technische Universität Dresden, Dresden; German Center for Diabetes Research (DZD), Munich, Neuherberg, Germany.

This PDF file contains:

- Figures S1 – S23
- Supplementary Information Text
- Supplementary References

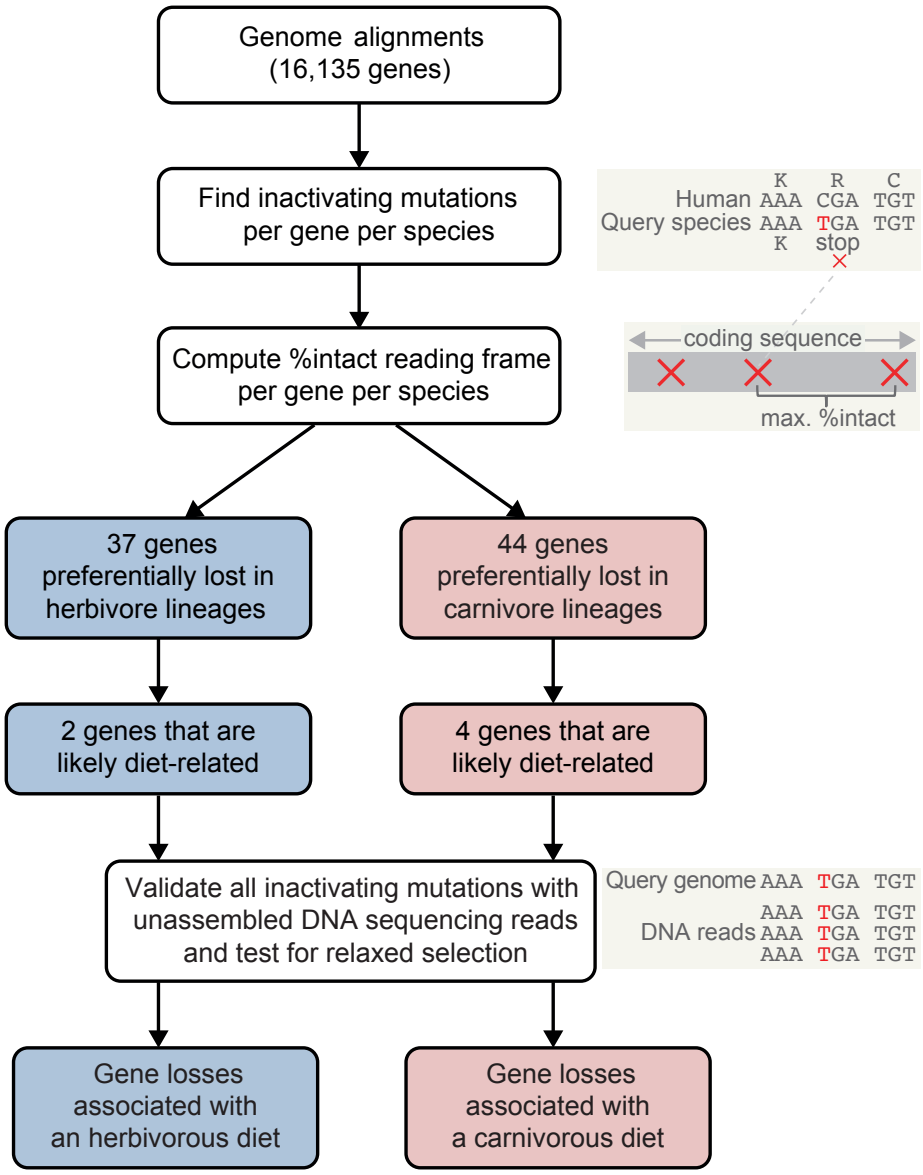

Figure S1: Workflow for identifying diet related convergent gene losses.
Insets illustrate inactivating mutations, the maximum percentage of the intact reading frame and validating mutations with DNA sequencing reads.

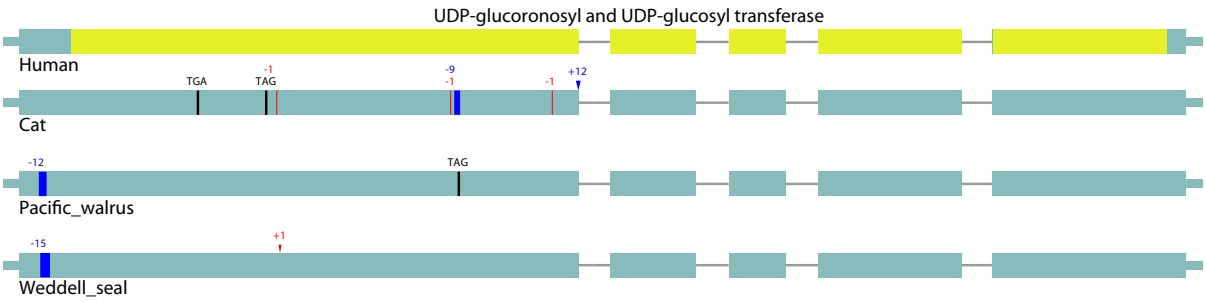

Figure S2: Inactivating mutations in *UGT1A6*.

General figure legend for Figures S2-S8, S17, S19-S22: Exons are represented by boxes proportional to their size. Introns are represented by horizontal lines. Small boxes to the left and right indicate the beginning and end of the gene. The exon-intron structure of the human gene is shown as a reference at the top. The location of functional domains in the human protein , downloaded from Ensembl Biomart (1), is indicated by colored boxes. The exon-intron structures shown below the human gene visualize the mutations that occurred in the orthologous gene in herbivores or carnivores. Here, a filled red box indicates an exon deletion and filled grey box missing genomic sequence. Vertical red lines show frameshifting deletions whereas vertical blue lines indicate frame preserving insertions or deletions. Arrow heads indicate frameshifting (red) or frame preserving (blue) insertions. The size of deletions or insertions is given on top of the mutation. Premature stop codons are indicated by black vertical lines and the corresponding triplet. Splice site mutations are shown by red letters at the end of an exon (donor mutation) or the beginning of an exon (acceptor mutation).

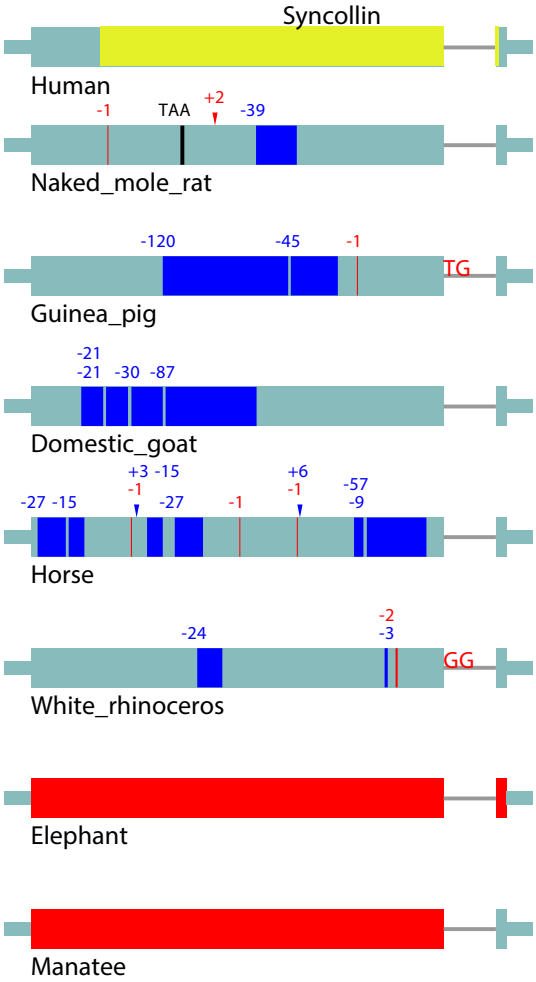

Figure S3: Inactivating mutations in *SYCN*. See Figure S2 for a description. The large deletion that occurred in the goat genome assembly of the breed 'Yunnan black' (NCBI GCA_000317765) is not found in the assembly of the breed 'San Clemente' (NCBI GCA_001704415), thus the loss of *SYCN* is likely not fixed in goats.

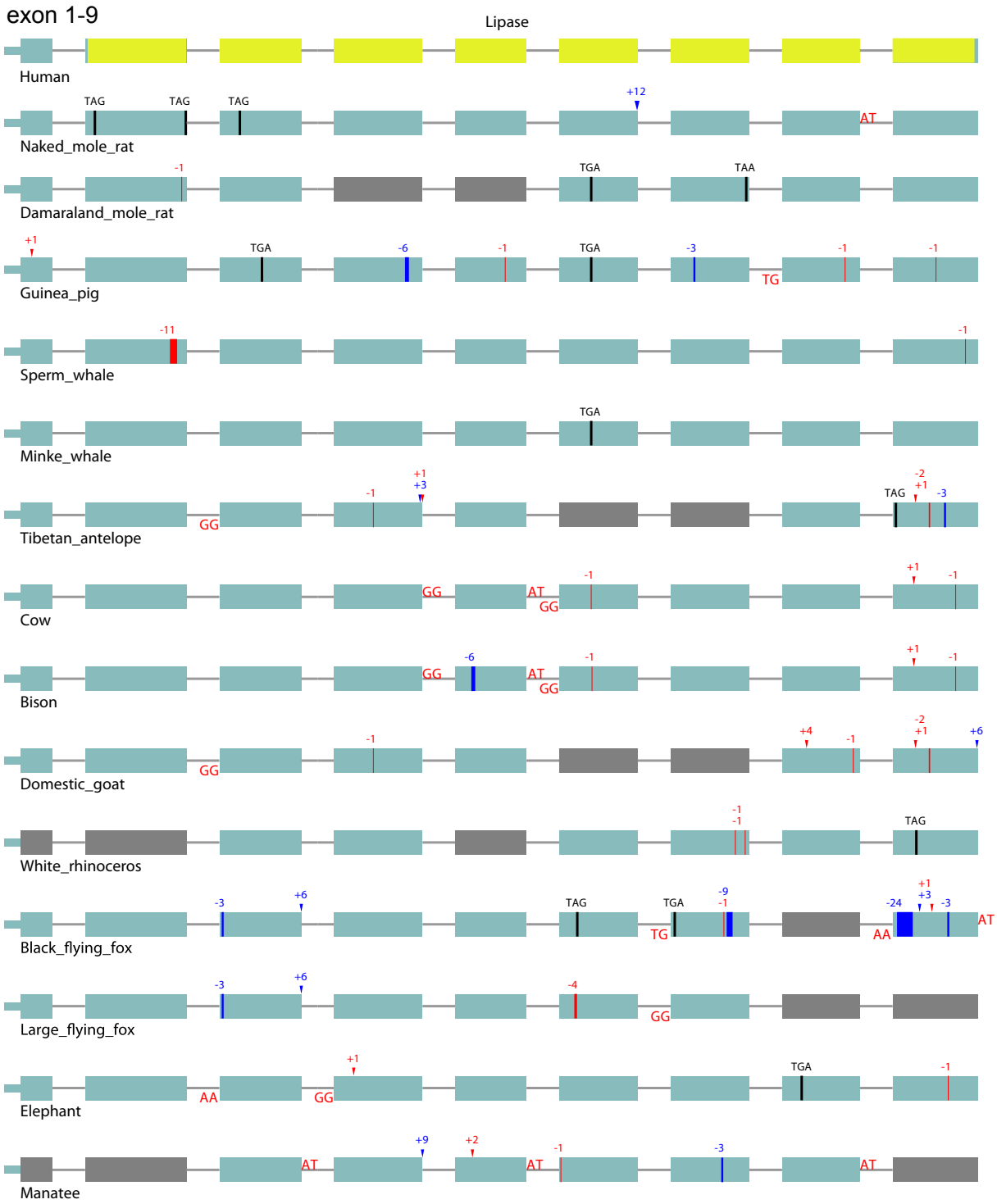

Figure S4: Inactivating mutations in *PNLIPRP1* (exons 1-9). See Figure S2 for a description. (Figure will be continued on next page.)

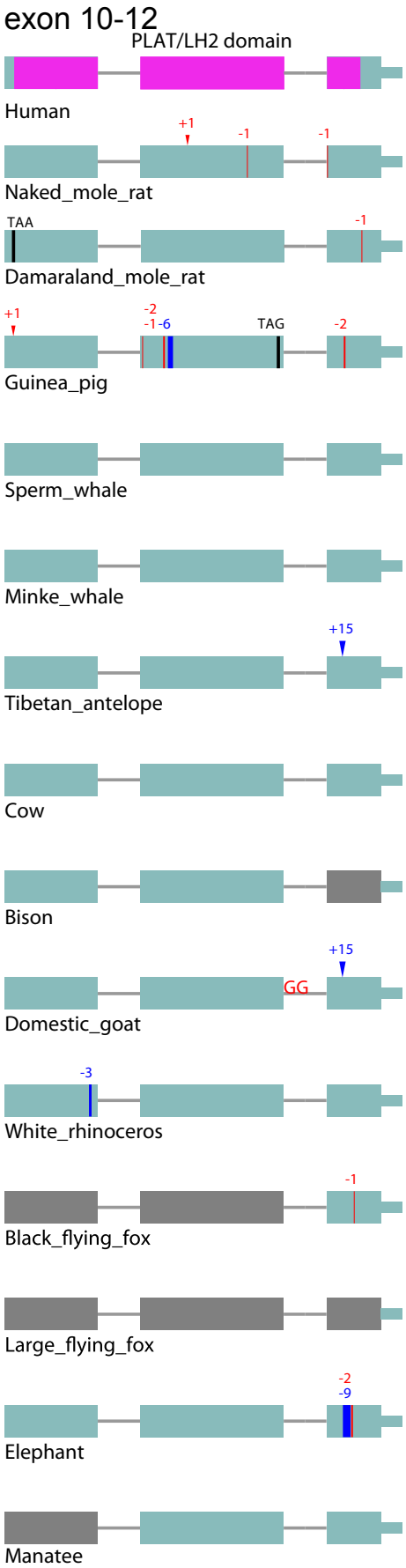

Figure S4 (continued from last page): Inactivating mutations in *PNLIPRP1* (exons 10-12). See Figure S2 for a description.

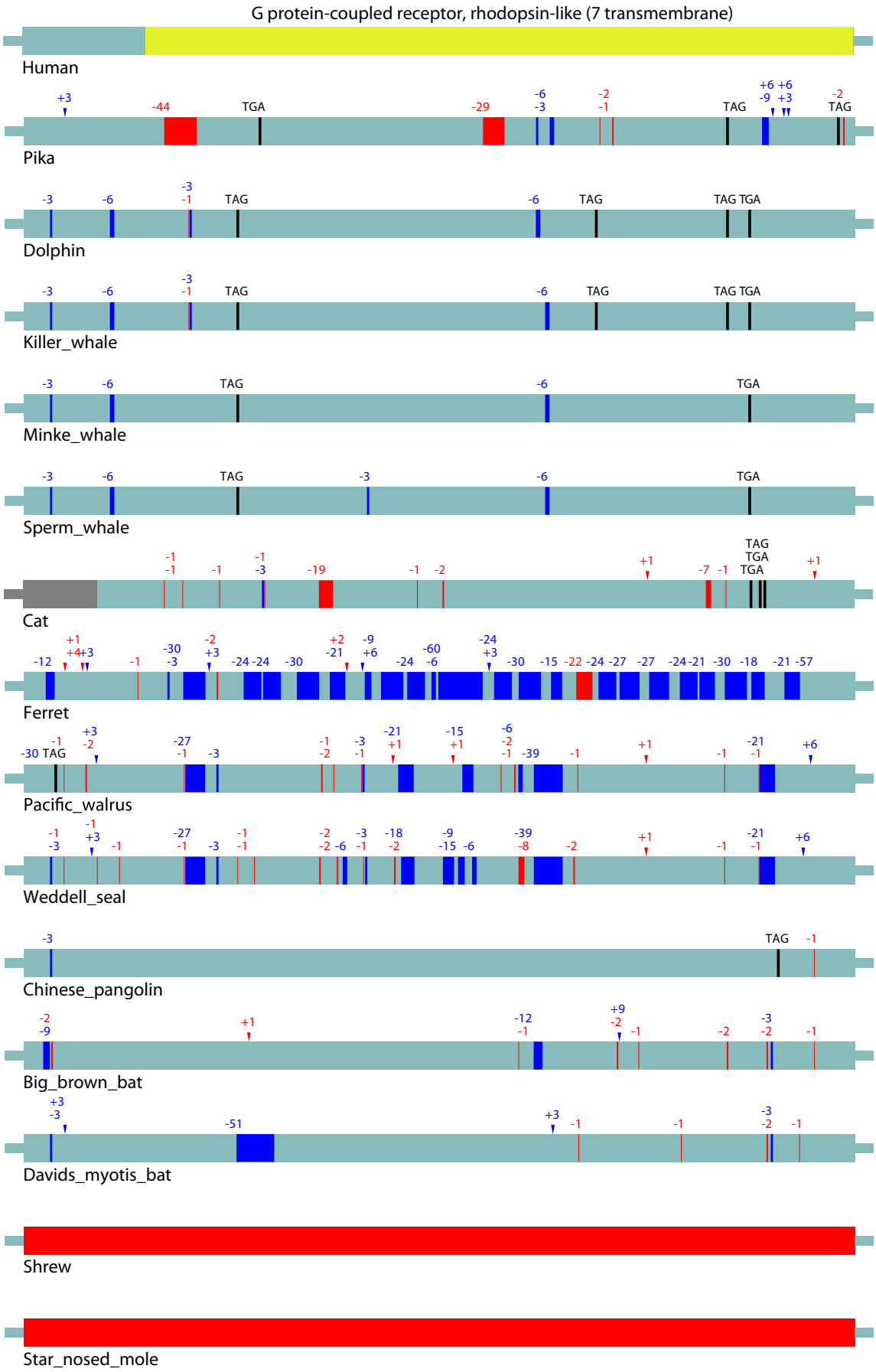

Figure S5: Inactivating mutations in *RXFP4*. See Figure S2 for a description.

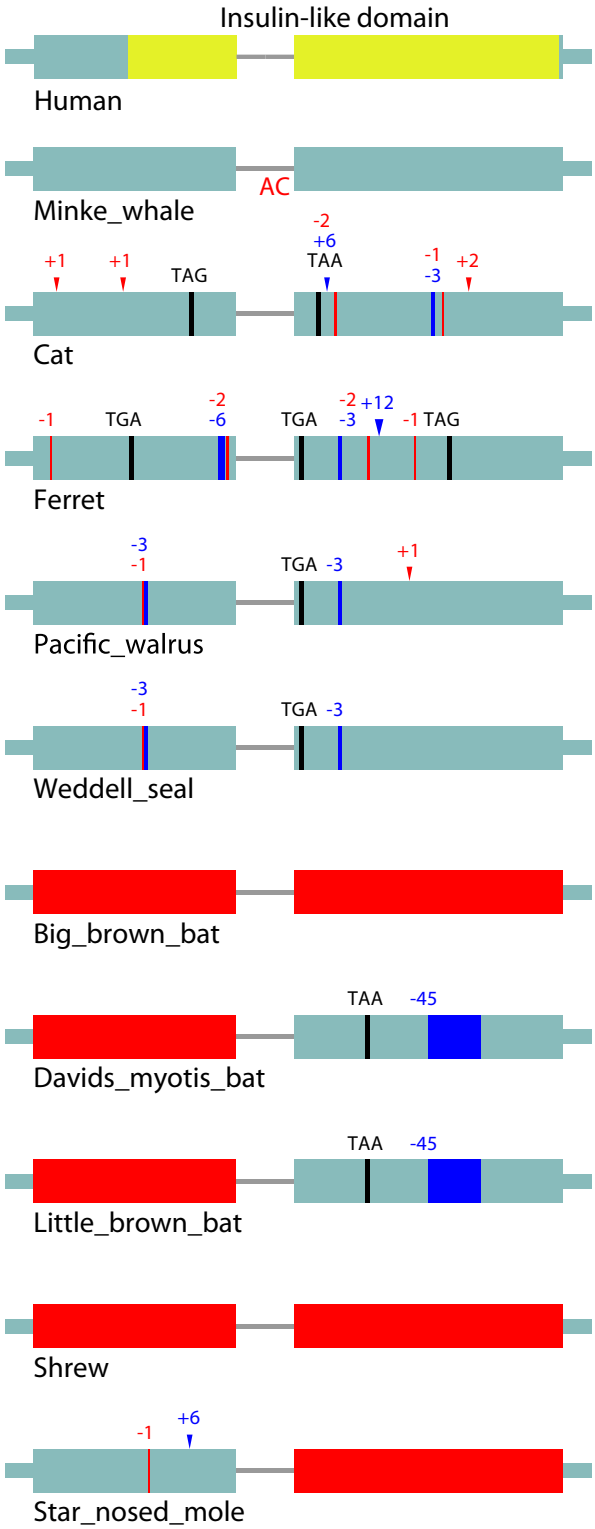

Figure S6: Inactivating mutations in *INSL5*. See Figure S2 for a description.

Figure S7: Inactivating mutations in *NOX1*. See Figure S2 for a description.

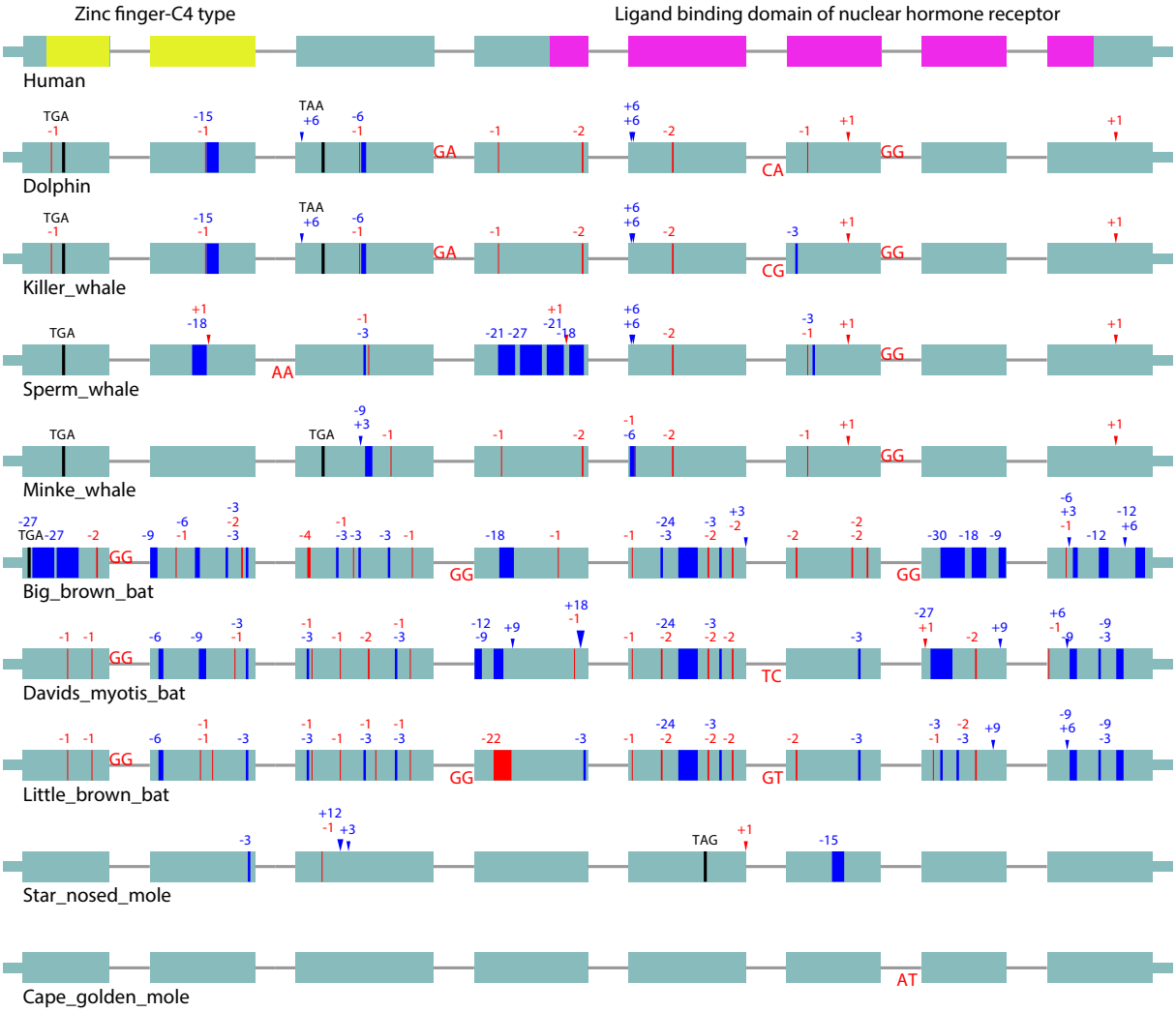

Figure S8: Inactivating mutations in *NR1I3*. See Figure S2 for a description. Please note that the acceptor mutation in the cape golden mole is likely polymorphic.

A) SYCN exon 1 stop codon in mole rat

```
Human           CCCTACTATGAGAAC
Ferret          CCCTACTACCAGAAC
Naked mole rat  CCCTACTAAGACAAC
Damara mole rat CCCTACTAAGACAAC
```

B) SYCN exon 1 frame-shift in guinea pig

```
Human                TCTGCCGGCACCT
Ferret               TCGGCGGGCGCCT
Guinea pig           ACCTCA-GCGCCT
Brazilian guinea pig ACCTCA-GCGCCT
```

C) SYCN exon 1 frame-shift in horse I

```
Human  GACCTGCCCTACCTG
Ferret GACCTGCCCTTCCTG
Horse  CACCTGCC-TTCCCG
Ass    CACCTGCC-TTCCCG
```

D) SYCN exon 1 frame-shift in horse II

```
Human  CCGCGC-----TGCGAG
Ferret CCGCGC-----TGCGAG
Horse  CCGCGCACGCACACGAG
Ass    CCGCGCACGCACACGAG
```

E) SYCN exon 1 frame-shift in rhinoceros

```
Human                CGCCGGGGCATCTTA
Ferret               CGCAAGGGTATCTTT
White rhinoceros     CGCCCG--GCTCCTC
Sumatran rhinoceros  CGCCCG--GCTCCTC
```

F) SYCN gene deletion in elephant

```
Human            taagtcctgatc...[    SYCN gene, ~3700 bp    ]...acagcctgcaac
African elephant taaggcctg---...[large deletion, 414 bp left]...---gtctgcatc
Asian elephant   taaggcctg---...[large deletion, 414 bp left]...---gtctgcatc
```

Figure S9: Inactivating mutations in *SYCN* that are shared between closely-related species.

Species that exhibit inactivating mutations are shown in red font. An analysis of sequencing reads of the Asian elephant, which does not have an assembled genome so far, indicates that the large deletion that removed the entire *SYCN* gene is shared between the African and Asian elephant with identical breakpoints. The human (hg38 assembly) coordinates are: (A) chr19:39204102-39204116, (B) chr19:39203931-39203946, (C) chr19:39204048-39204062, (D) chr19:39203994-39204005, (E) chr19:39203898-39203912, (F) chr19:39201882-39205605.

A) PNLIPRP1 exon 9 frame-shift in guinea pig

```
Human                TTTGCTGGCAGGACA
Ferret               TTTGCCGGCAAAACA
Guinea pig           TTTGTT-GTAAGACA
Brazilian guinea pig TTTGTT-GTAAGACA
```

B) PNLIPRP1 exon 11 stop codon in guinea pig

```
Human                GAAGAGAAGACA
Ferret               GATGAGAAGACA
Guinea pig           CAGGAATAGACA
Brazilian guinea pig CAGGAATAGACA
```

C) PNLIPRP1 exon 4 frame-shift in goat and antelope

```
Human            TCCCAAGCCACCTAC
Killer whale     TCCCAGACCACCTAC
Sheep            TCCCAGA-CACTTAC
Bighorn sheep    TCCCAGA-CACTTAC
Goat             TCCCAGA-CACTTAC
Tibetan antelope TCCCAGA-CACTTAC
```

D) PNLIPRP1 exon 9 frame-shift in Bovidae

```
Human            CAAGGATG-CCCACAG
Killer whale     GAAGGGTG-CCCGCAG
Cow              GAAGGGTGTCCCACAG
Zebu cattle      GAAGGGTGTCCCACAG
Bison            GAAGGGTGTCCCACAG
Wild yak         GAAGGGTGTCCCACAG
Water buffalo    GAAGGGTGTCCCACAG
Sheep            GAAGGGTGTCCCGCAG
Bighorn sheep    GAAGGGTGTCCCGCAG
Goat             GAAGGGTGTCCCGCAG
Tibetan antelope GAAGGGTGTCCCGCAG
```

E) PNLIPRP1 exon 7 splice-site mutation in fruit bats

```
Human            tcagGTTTT
Big brown bat    tcagGTTTT
Black flying fox tctgGTTTT
Large flying fox tctgGATTT
```

F) PNLIPRP1 exon 7 frame-shift in fruit bats

```
Human            AATGCCCTGTCTCAGATCGTGGAT
Big brown bat    AACGCCTTGTCCCAGATTGTGGAC
Black flying fox AATGCC----------TTATGGAC
Large flying fox AATGCC----------TTATGGAC
Egyptian rousette AATATC----------CTTTCGAC
```

G) PNLIPRP1 exon 4 frame-shift in elephant

```
Human            GTGAACTG-CATCTGC
Cape golden mole GTGAACTG-CATCTGC
African elephant GTGAACTGTTATCTGT
Asian elephant   GTGAACTGTTATCTGT
```

H) PNLIPRP1 exon 8 stop codon in elephant

```
Human            AATCACCTAAGAAGC
Cape golden mole AATCACTTAAGAAGC
African elephant AATCACTGAAGAAGC
Asian elephant   AATCACTGAAGAAGC
```

I) PNLIPRP1 exon 6 stop codon in baleen whales

```
Human                  GAGGTGCGACTTGAT
Bottlenose dolphin     GAGGTCCGACTCGAT
Killer whale           GAGGTCCGACTCGAT
Beluga whale           GAGGTCCGACTCGAT
Yangtze river dolphin  GAGGTCCGACTTGAT
Sperm whale            GAGGTCCGACTCGAT
Minke whale            GAGGTCTGACTCGAT
Bowhead whale          GAGGTCTGACTTGAT
```

J) PNLIPRP1 exon 7 frame-shift in rhinoceros

```
Human                            ATCTGGGCGGgt
Ferret                           ATTTGGGAAGgt
White rhinoceros                 ATCTGA-CAGgt
Sumatran rhinoceros (copy 1)     GTCTGC-CAGgt
Sumatran rhinoceros (copy 2)     ATCTGG-CAGgt
```

Figure S10: Inactivating mutations in *PNLIPRP1* that are shared between closely-related species.

Species that exhibit inactivating mutations are shown in red font. The Sumatran rhinoceros exhibits a duplication of exon 7; however, both copies exhibit the same frameshift found in the White rhinoceros. The human (hg38 assembly) coordinates are:
(A) chr10:116601132-116601146, (B) chr10:116605540-116605551,
(C) chr10:116594784-116594798, (D) chr10:116601099-116601110,
(E) chr10:116598043-116598051, (F) chr10:116598124-116598147,
(G) chr10:116594751-116594765, (H) chr10:116600070-116600084,
(I) chr10:116597869-116597883, (J) chr10:116598157-116598165.

A) RXFP4 exon 1 stop codon in cetaceans

```
Human                  CCCTTTTGGGCAGCC
Cow                    CCCTTCTGGGCAGCC
Bottlenose dolphin     CCCTTCTAGGCAGCC
Killer whale           CCCTTCTAGGCAGCC
Beluga whale           CCCTTCTAGGCAGCC
Yangtze river dolphin  CCCTTCTAGGCAGCC
Sperm whale            CCCTTCTAGGCAGCC
Minke whale            CCCTTCTAGGCAGCC
Bowhead whale          CCCTTCTAGGCAGCC
```

B) RXFP4 exon 1 frame-shift in cat and pinnipeds

```
Human              CTCATCACAGCGCTGAGC
Horse              CTCATCACGGCGCTGAGT
Cat                CTCACCCC--CGGCGCAT
Leopard            CTCATCCC--CGGCGCAT
Pacific walrus     CTCACTCC--CGCCAAGT
Weddell seal       CTCACTCC--GGCTGAGT
Hawaiian monk seal CTCACTCC--GGCTGAGT
```

C) RXFP4 exon 1 frame-shift in cat, ferret and pinnipeds

```
Human              CGGGAGCCCCGGCAGGCT
Horse              CGGGAGCCACGAAGGGCC
Cat                CAGGAGCC-TGATGGGCC
Cheetah            CAGGAGCC-TGATGGGCC
Amur tiger         CAGGAGCC-TGATGGGCC
Leopard            CAGGAGCC-TGATGGGCC
Ferret             CAGGAGCC-TGATGGGGC
Sea otter          CAGGAGCC-TGATGGGGC
Pacific walrus     CAGGAGCC-TGATGGACC
Weddell seal       CAGGAGCC-TGACGGGCC
Hawaiian monk seal CAGGAGCC-TGACGGGCC
```

D) RXFP4 exon 1 frame-shift in carnivorous bats

```
Human              CAGGGCGGAGGCTGGGTG
Black flying fox   CCGGGCCGCGGCTGGGTG
Large flying fox   CCGGGCCGCGGCTGGGTG
Big brown bat      CAGAAC-----CGGGGTG
David's myotis     CAGGAC-----CGGGGTG
Brandt's myotis    CAGAAC-----CGGGGTG
Little brown bat   CAGAAC-----CGGGGTG
```

Figure S11: Inactivating mutations in *RXFP4* that are shared between closely-related species.

Species that exhibit inactivating mutations are shown in red font. Please note that the carnivorous bats all share the same 5 bp deletion, which may indicate a shared loss of *RXFP4* that already occurred in their common ancestor. However, this mutation is downstream of the last transmembrane domain and rather close to the C-terminus. Therefore, we are conservative and do not count this mutation as evidence for a shared gene loss in these bats in Figure 1. While this 5 bp deletion is the only potentially inactivating mutation in the little brown bat, the other 3 bats (Big brown bat, David's and Brandt's myotis) all have several other, lineage-specific inactivating mutations, which shows the *RXFP4* is clearly lost in these species. The human (hg38 assembly) coordinates are: (A) chr1:155941992-155942006, (B) chr1:155942100-155942117, (C) chr1:155942649-155942666, (D) chr1:155942709-155942726.

A) INSL5 exon 1 frame-shift in ferret

```
Human        TCCATTTTCACTCTG
Horse        GCCATTTTCACCTTA
Ferret       TCCCTTT-GACTTTA
Sea otter    TCCCTTT-GACTTTA
```

B) INSL5 exon 1 frame-shift in cat

```
Human        GAGTCTGT-GAGACTC
Horse        GAGTCCAG-GAAGCTC
Cat          GAGTCTATAGAAGCTC
Cheetah      GAGTCTATAGAAGCTC
Amur tiger   GAGTCTATAGAAGCTC
Leopard      GAGTCTATAGAAGCTC
```

C) INSL5 exon 2 stop codon in ferret and pinnipeds

```
Human              CTGAGACAGGAAAC
Horse              CTGAGTGGGACAAC
Ferret             CTGAGTGAGGAAAT
Sea otter          CTGAGTGAGGAAAT
Pacific walrus     CTGAGTGAGGAAAC
Weddell seal       CTGAGTGAGGAAAC
Hawaiian monk seal CTGAGTGAGGAAAC
```

D) INSL5 exon 2 stop codon in cat

```
Human        TCCTTCCAGCTCCCA
Horse        TACTTCCAGCTGCCA
Cat          TACTTTTAACTTCCA
Cheetah      TACTTCTAACTTCCA
Amur tiger   TACTTCTAACTTCCA
Leopard      TACTTCTAACTTCCA
```

E) INSL5 exon 2 stop codon in myotis bats

```
Human            CCAGCGCAAAACCTT
Black flying fox ATGGCACAAAACCTT
Large flying fox ATGGCACAAAACCTT
David's myotis   ACACCGTAAAACCTT
Brandt's myotis  ACACCGTAAAACCTT
Little brown bat ACACCGTAAAACCTT
```

Figure S12: Inactivating mutations in *INSL5* that are shared between closely-related species.

Species that exhibit inactivating mutations are shown in red font. (E) It should be noted that, in addition to the shared stop codon, the deletion of *INSL5* exon 1 in the three myotis bats has the same breakpoint and is therefore shared between these species. The human (hg38 assembly) coordinates are: (A) chr1:66801198-66801212, (B) chr1:66801138-66801152, (C) chr1:66798232-66798245, (D) chr1:66798217-66798231, (E) chr1:66798175-66798189.

A) NOX1 exon 2 frame-shift in cetaceans

Human	AAGGCCGACAAATAC
Cow	AAGGCCACCAAGTAC
Bottlenose dolphin	AAGGTAAC-AAGTAC
Killer whale	AAGGTAAC-AAGTAC
Beluga whale	AAGGTAAC-AAGTAC
Yangtze river dolphin	AAGGTAAC-AAGTAC
Sperm whale	AAGGCTAC-AAGTAC
Minke whale	AAGGCAAC-AAGTAC
Bowhead whale	AAGGCAAC-AAGTAC

B) NOX1 exon 9 stop codon in cetaceans

Human	CATATCCGAGCAGCA
Cow	CATATCCGAGCAGTG
Bottlenose dolphin	CATATCTGAGCAGTG
Killer whale	CATATCTGAGCAGTG
Beluga whale	CATATCTGAGCAGTG
Yangtze river dolphin	CATATCTGAGCAGTG
Sperm whale	CATATCTGAGCAGTG
Minke whale	CATATCTGAACAGTG
Bowhead whale	CATATCTGAACAGTG

C) NOX1 exon 3 stop codon in pangolins

Human	TGTGCCCGAGCGTCT
Horse	TGGGCCCGAGCCTCT
Chinese pangolin	TGGGTCTGAGCCTCT
Malayan pangolin	TGGGTCTGAGCCTCT

D) NOX1 exon 9 stop codon in pangolins

Human	CATATCCGAGCAGCA
Horse	CATATCCGAGCAGCA
Chinese pangolin	CATATCTGAGCAGTG
Malayan pangolin	CATATCTGAGCAGTG

Figure S13: Inactivating mutations in *NOX1* that are shared between closely-related species.

Species that exhibit inactivating mutations are shown in red font. The human (hg38 assembly) coordinates are: (A) chrX:100870743-100870757, (B) chrX:100850210-100850224, (C) chrX:100863569-100863583, (D) chrX:100850210-100850224.

A) NR1I3 exon 1 stop codon in cetaceans

Human	GGGGACCAAGCCACA
Cow	GGGGACCGGGCTACA
Bottlenose dolphin	AGGAACTGAGCCACA
Killer whale	AGGAACTGAGCCACA
Beluga whale	AGGAACTGAGCCACA
Yangtze river dolphin	AGGAACTGAGCCACA
Sperm whale	AGGAACTGAGCCACA
Minke whale	AGGCACTGAGCCACA
Bowhead whale	AGGCACTGAGCCACG

B) NR1I3 exon 5 frame-shift in cetaceans

Human	GTGGAAATCTGTCACATC
Cow	ATAGAAATCTGTCATATT
Bottlenose dolphin	GTAGAAAT--GTCACATT
Killer whale	GTAGAAAT--GTCACATT
Beluga whale	GTAGAAAT--GTCACATT
Yangtze river dolphin	GTAGAAAT--GTCACATT
Sperm whale	GTAGAAAT--GTCACATT
Minke whale	GTAGAAAT--GTCACAAT
Bowhead whale	GTAGAAAT--GTCACATT

C) NR1I3 exon 2 frame-shift in carnivorous bats

Human	GTCAGCAAGACTCAG
Black flying fox	ATCAGCAAGGCCCAG
Large flying fox	ATCAGCAAGGCCCAG
Big brown bat	GTCAGCG-GGCCCGG
David's myotis	GTCAGCA-GGCCCAG
Brandt's myotis	GTCAACA-GGCCCAG
Little brown bat	GTCAACA-GGCCCAG

D) NR1I3 exon 5 frame-shift in carnivorous bats

Human	TACACAATTGAAGAT
Black flying fox	TACACAGAAGAAGAC
Large flying fox	TACACAGAAGAAGAC
Big brown bat	TACCCGA--GAAGAC
David's myotis	TACCCGA--GAAGAT
Brandt's myotis	TGCCCGA--GAAGAC
Little brown bat	TGCCCGA--GAAGAC

Figure S14: Inactivating mutations in *NR1I3* that are shared between closely-related species.

Species that exhibit inactivating mutations are shown in red font. The human (hg38 assembly) coordinates are: (A) chr1:161236509-161236523, (B) chr1:161231411-161231428, (C) chr1:161235908-161235922, (D) chr1:161231339-161231353.

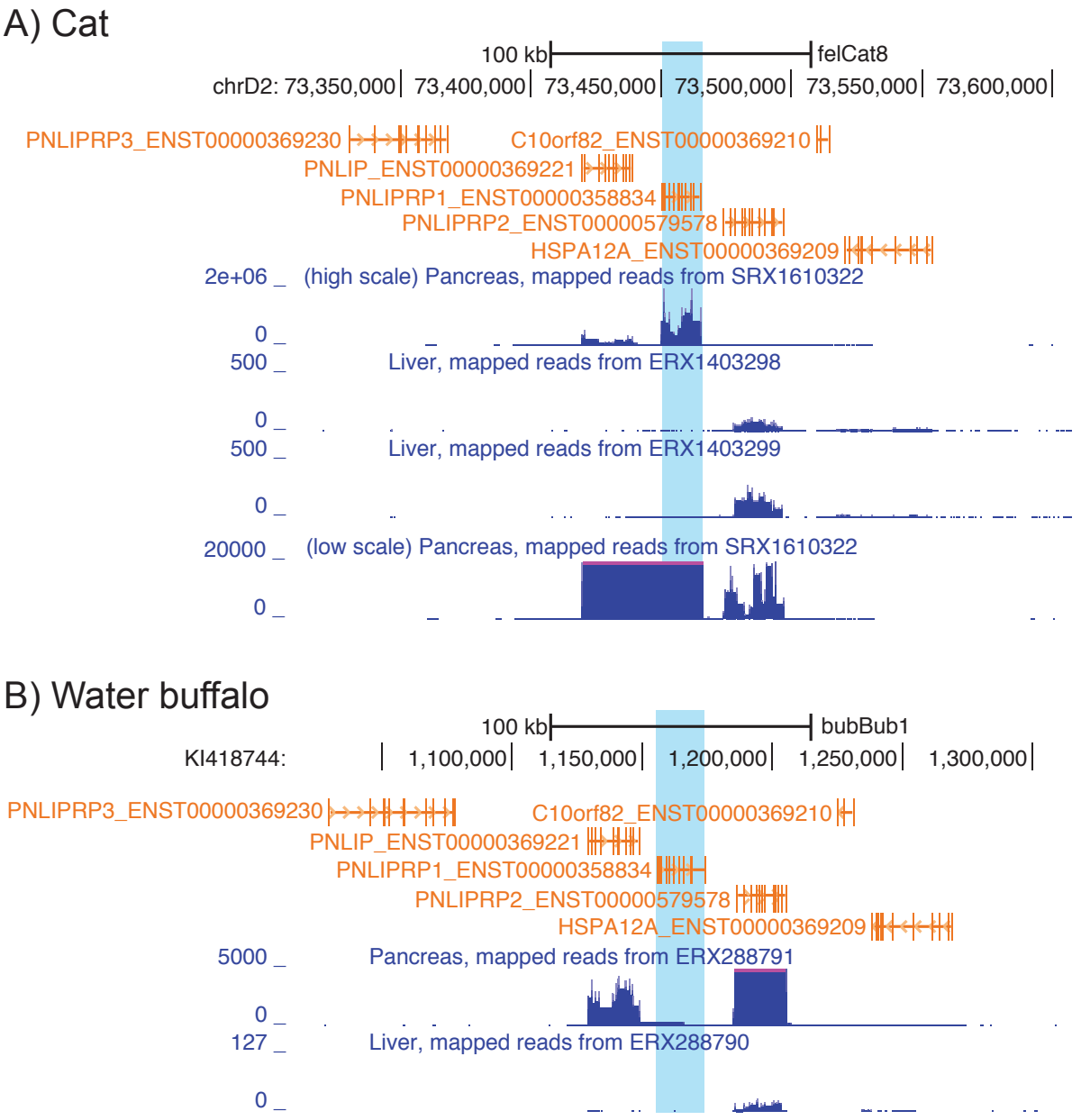

Figure S15: Expression of *PNLIPRP1* in the pancreas of Water buffalo and cat.
UCSC genome browser (2) screenshots show in orange human transcripts mapped with CESAR 2.0 (3) and in blue mRNA expression data (tissue and data source identifier is indicated).
(A) As a control, we confirmed that *PNLIPRP1* (highlighted in cyan) is expressed in the pancreas of the cat, a carnivorous species with an intact gene. Since *PNLIPRP1* appears to be expressed at a very high level, we also visualized pancreas expression using a different Y-axis scaling (bottom of this panel) to show that *PNLIPRP2* is also expressed in cat.
(B) In contrast to the cat, *PNLIPRP1* is not expressed anymore in the pancreas of the buffalo, an herbivore that lost this gene, while the neighboring *PNLIP* and *PNLIPRP2* genes are expressed in the pancreas.

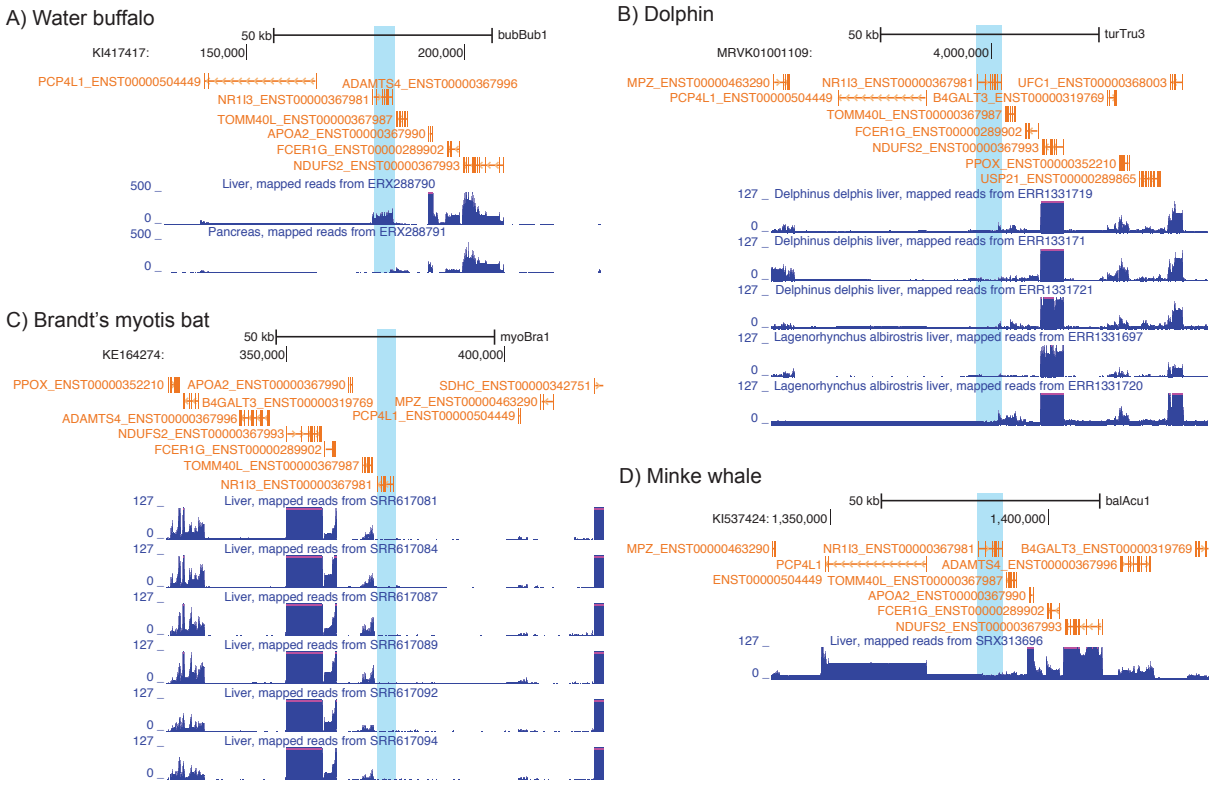

Figure S16: Expression of *NR1I3* in the liver of Water buffalo, cetaceans and Brandt's myotis bat.

UCSC genome browser (2) screenshots show in orange human transcripts mapped with CESAR 2.0 (3) and in blue mRNA expression data (tissue and data source identifier is indicated).

(A) As a control, we confirmed that *NR1I3* (highlighted in cyan) is expressed in the liver of the buffalo, an herbivore with an intact gene.

(B-D) *NR1I3* has no relevant exon expression or is not expressed at all anymore in the liver of dolphins (B), Brandt's myotis bat (C) and the minke whale (D). In (B) we mapped available liver mRNA expression data from the two dolphins *Delphinus delphis* and *Lagenorhynchus albirostris* to the genome of the bottlenose dolphin.

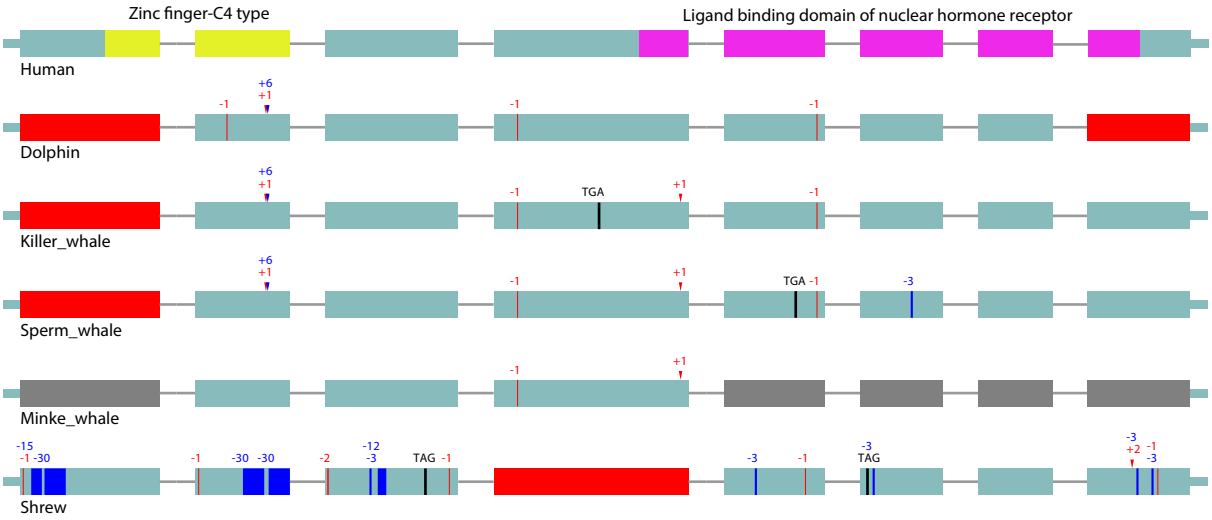

Figure S17: Inactivating mutations in *NR1I2*. See Figure S2 for a description.

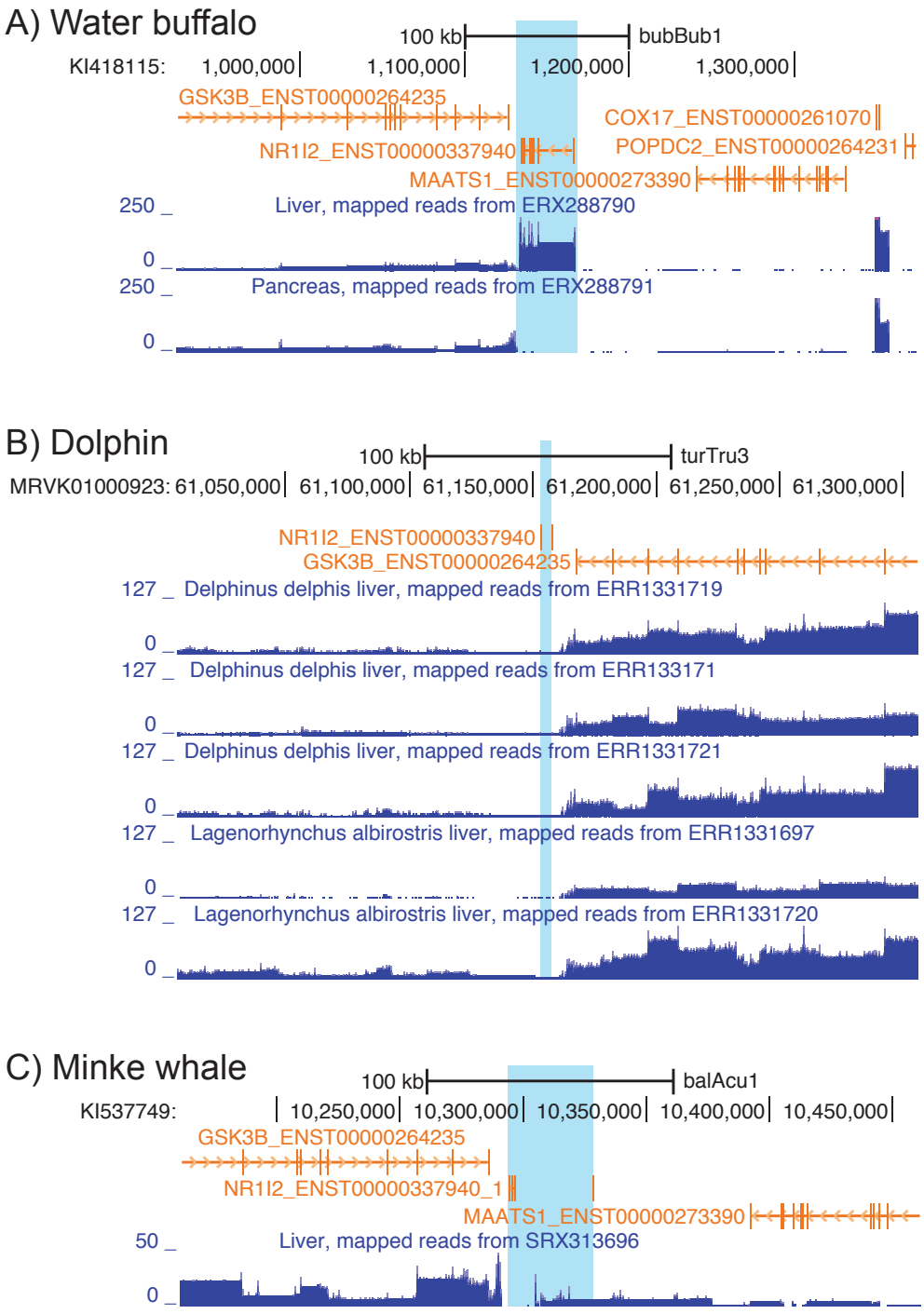

Figure S18: Expression of *NR1I2* in the liver of Water buffalo and cetaceans.
UCSC genome browser (2) screenshots show in orange human transcripts mapped with CESAR 2.0 (3) and in blue mRNA expression data (tissue and data source identifier is indicated).
(A) Like *NR1I3*, *NR1I2* (highlighted in cyan) is expressed in the liver of the buffalo, an herbivore with an intact gene.
(B-C) The remnants of *NR1I2* have no relevant exon expression in the liver of dolphins (B) and the minke whale (C). In (B) we mapped available liver mRNA expression data from the two dolphins *Delphinus delphis* and *Lagenorhynchus albirostris* to the genome of the bottlenose dolphin.

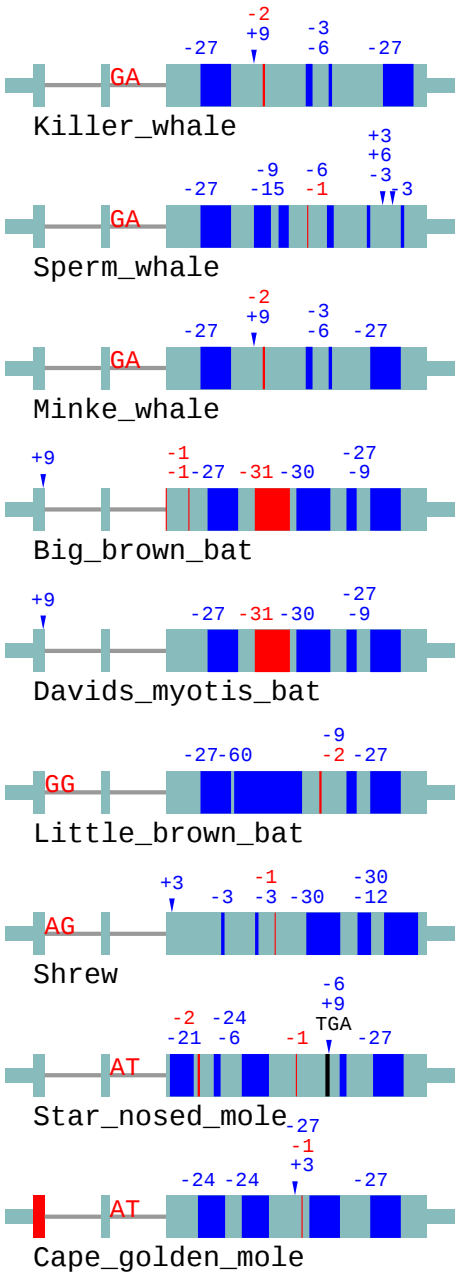

Figure S19: Inactivating mutations in *PHGR1*. See Figure S2 for a description. This gene has no annotated protein domains.

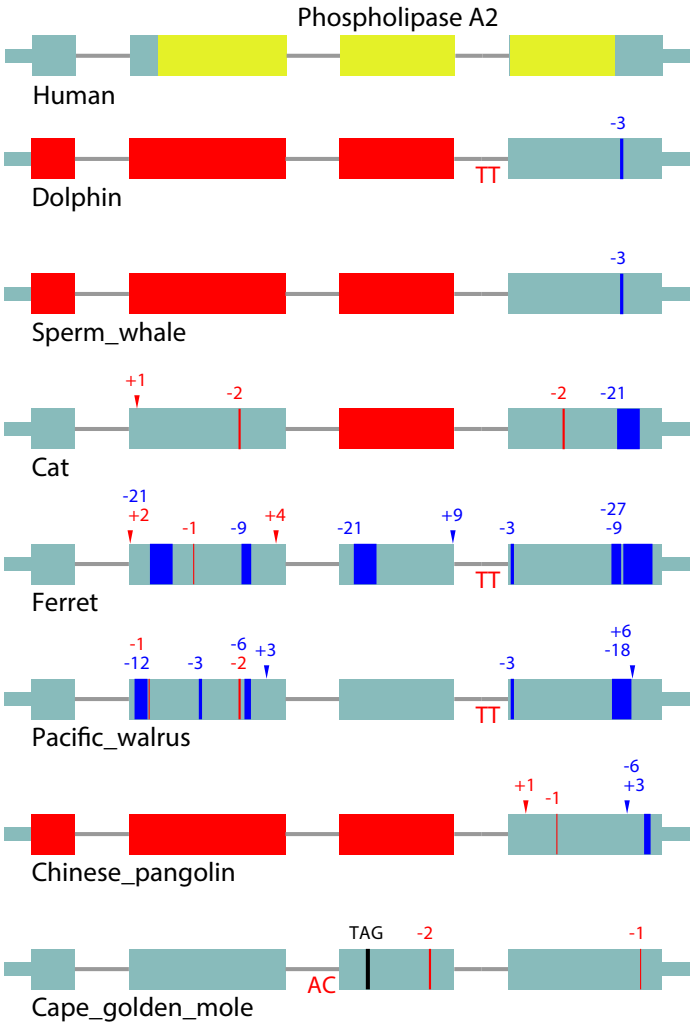

Figure S20: Inactivating mutations in *PLA2G2A*. See Figure S2 for a description.

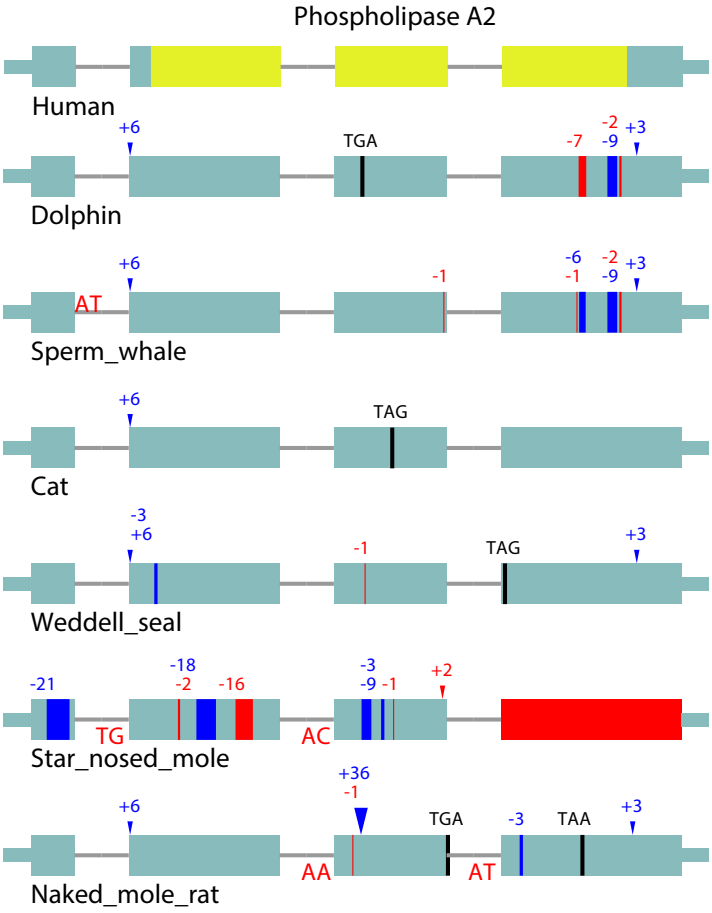

Figure S21: Inactivating mutations in *PLA2G2C*. See Figure S2 for a description.

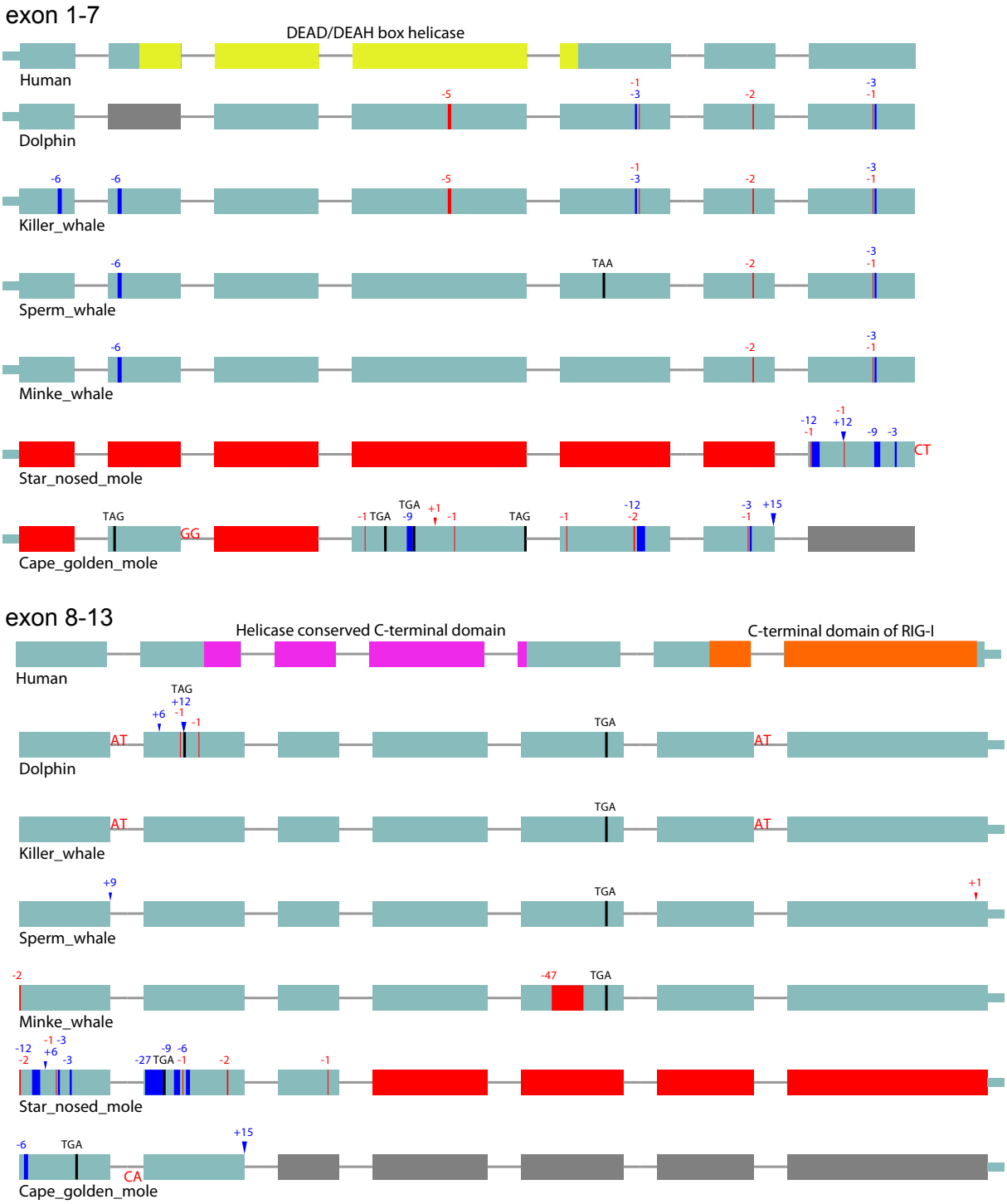

Figure S22: Inactivating mutations in *DDX58*. See Figure S2 for a description.

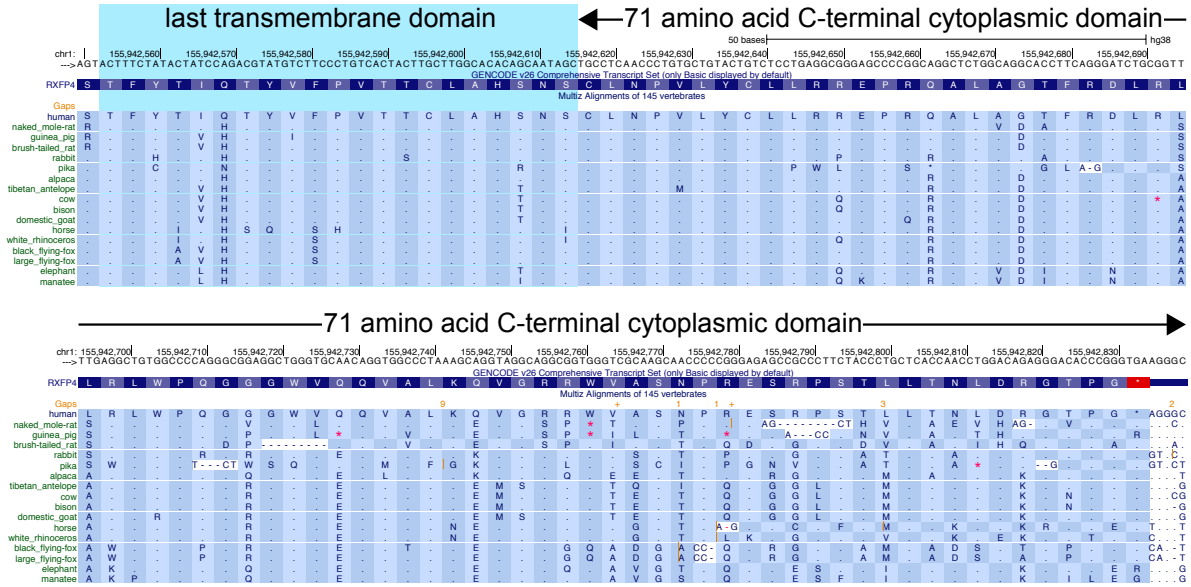

Figure S23: Evolutionary variability of the RXFP4 C-terminal domain in herbivores. UCSC genome browser screenshots show the C-terminus of RXFP4 in obligate herbivores. The last transmembrane domain is highlighted in cyan, which is followed by the 71 amino acid long C-terminal cytoplasmic domain. *RXFP4* is truly inactivated in the herbivorous pika, where the gene exhibits several inactivating mutations spread across the gene (Figure S5, Datasets S3, S7). In contrast to the pika, other herbivores only have inactivating mutations downstream of the last transmembrane domain, as shown here. Thus, these mutations solely shorten the 71 amino acid long C-terminal cytoplasmic domain. Given that many conserved genes have inactivating mutations close to the C-terminus (4, 5), these mutations are therefore no indication of gene loss. Stop codons are indicated by red asterisks.

Supplementary Information Text

Materials and Methods

Classification of placental mammals into obligate herbivores and carnivores

We applied a rather strict definition of herbivory and carnivory. For instance, we excluded the mainly carnivorous polar bears that are known to eat kelp and the giant panda bears that despite their bamboo-dominated diet also prey on rodents occasionally. Similarly, we excluded primates that have primarily an herbivorous diet but also feed on insects. This strict classification resulted in six independent herbivore lineages and five independent carnivore lineages (Fig. 1).

Gene ontology enrichment

To assess whether the sets of convergently lost genes share common biological functions, we performed an enrichment analysis. We downloaded human gene – ontology term associations from the Gene Ontology (6, 7). Importantly, not every human gene could have been detected by our screen for convergent gene loss, since some genes do not align to a sufficiently large number of mammals. Therefore, to avoid any bias, we restricted our enrichment analysis to those genes for which ≥50% of the herbivore and ≥50% of the carnivore species had aligning sequence and thus convergent loss could have been detected. This resulted in a total of 14,923 genes that were annotated in at least one GO term. We only tested GO terms that are annotated for at least five genes in this entire set. Separately for herbivores and carnivores, we then performed hypergeometric tests using the R function ‘phyper’ (8) to assess if convergently-lost genes are statistically enriched in a particular GO term annotation. The resulting p-values were then adjusted with a Holm-Bonferroni correction for multiple testing using the R implementation ‘p.adjust’ (9).

Selection rate analysis

To corroborate convergent gene loss in herbivores, we used RELAX as implemented in the HYPHY package (version 2.3.720171127beta) (10, 11). We tested whether the branches corresponding to herbivore lineages that lost a gene exhibit relaxed selection rates compared to those branches corresponding to carnivore lineages that possess an intact gene. For this purpose, we determined the common ancestors of the gene loss herbivores in the phylogeny and labelled all branches in the respective subtrees as foreground. An analogous approach was applied to test for relaxed selection in gene loss carnivore branches. Input sequences were obtained from pair-wise CESAR alignments (5). Multiple sequence alignments were computed with PRANK using the species phylogeny (parameter -once) and keeping gaps based on phylogenetic signal (parameter -F) (12). These alignments were then used as input for RELAX. We inferred relaxed selection if the relaxation parameter k is < 1 and the p-value is < 0.05.

Analysis of expression data

As an additional line of evidence, we investigated gene expression data of species that exhibit gene loss. It should be noted that pancreas expression data of *SYCN*-loss species and expression data of colon tissue, where *RXFP4*, *INSL5* and *NOX1* are specifically expressed, was not available. However, pancreas and liver expression data were available, both for species that lost *PNLIPRP1*, *NR1I3* or *NR1I2*, and for species that have intact genes (serving as a control to verify expression of these pancreas- or liver-specific genes). We downloaded from the Sequence Read Archive (SRA) (13) gene expression data of Water buffalo pancreas and liver tissue (SRA study ERP003627, ref. (14)), cat pancreas (SRA study SRP071078, provided by 99 Lives Cat Genome Sequencing Initiative), liver of cat and two dolphin species (*Delphinus delphis* and *Lagenorhynchus albirostris*, SRA study ERP014610, ref. (15)), and Brandt's myotis bat liver (SRA study SRP017183, ref. (16)). Raw read data was processed with fastq-dump using parameters for removing technical reads (skip-technical), filtering low-quality reads (read-filter pass), and removing tags (clip). Afterwards, reads were mapped to the respective genomes with STAR (version 2.4.2a) (17). For computing genomic indices, we varied the number of bins (genomeChrBinNbits) according to the number of scaffolds and total sequence lengths of the assemblies. Reads were mapped with recommended defaults as specified in the STAR manual (maximum ratio of mismatches (outFilterMismatchNoverLmax) of 0.04 and maximum of 20 mappings per read (outFilterMultimapNmax)). The coverage of each base in the genomes by mapped reads was computed with BEDtools (18).

Supplementary References

1. Kinsella RJ*, et al.* (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database : the journal of biological databases and curation* 2011:bar030.
2. Casper J*, et al.* (2018) The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res* 46(D1):D762-D769.
3. Sharma V, Schwede P, & Hiller M (2017) CESAR 2.0 substantially improves speed and accuracy of comparative gene annotation. *Bioinformatics* 33(24):3985-3987.
4. MacArthur DG*, et al.* (2012) A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335(6070):823-828.
5. Sharma V, Elghafari A, & Hiller M (2016) Coding exon-structure aware realigner (CESAR) utilizes genome alignments for accurate comparative gene annotation. *Nucleic Acids Res* 44(11):e103.
6. Ashburner M*, et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25(1):25-29.
7. The Gene Ontology Consortium (2017) Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res* 45(D1):D331-D338.
8. Johnson NL, Kemp AW, & Kotz S (2005) *Univariate discrete distributions* (John Wiley & Sons, Hoboken, New Jersey, USA).
9. Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* 6:65-70.
10. Pond SL, Frost SD, & Muse SV (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21(5):676-679.
11. Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, & Scheffler K (2015) RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular biology and evolution* 32(3):820-832.
12. Loytynoja A & Goldman N (2008) Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320(5883):1632-1635.
13. Kodama Y, Shumway M, Leinonen R, & International Nucleotide Sequence Database C (2012) The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res* 40(Database issue):D54-56.
14. Williams JL*, et al.* (2017) Genome assembly and transcriptome resource for river buffalo, Bubalus bubalis (2n = 50). *Gigascience* 6(10):1-6.
15. Berthelot C, Villar D, Horvath JE, Odom DT, & Flicek P (2018) Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nat Ecol Evol* 2(1):152-163.
16. Seim I*, et al.* (2013) Genome analysis reveals insights into physiology and longevity of the Brandt's bat Myotis brandtii. *Nature communications* 4:2212.
17. Dobin A*, et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15-21.
18. Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841-842.