

# The Imprinted *NPAP1* Gene in the Prader–Willi Syndrome Region Belongs to a *POM121*-Related Family of Retrogenes

Lisa C. Neumann<sup>1,3</sup>, Nathalie Feiner<sup>2,4</sup>, Axel Meyer<sup>2</sup>, Karin Buiting<sup>1</sup>, and Bernhard Horsthemke<sup>1,\*</sup>

<sup>1</sup>Institut für Humangenetik, Universitätsklinikum Essen, Universität Duisburg-Essen, Germany

<sup>2</sup>Laboratory for Zoology and Evolutionary Biology, Department of Biology, University of Konstanz, Germany

<sup>3</sup>Present address: Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom

<sup>4</sup>Present address: Department of Zoology, University of Oxford, Oxford, United Kingdom

\*Corresponding author: E-mail: [bernhard.horsthemke@uni-due.de](mailto:bernhard.horsthemke@uni-due.de).

Accepted: January 23, 2014

**Data deposition:** The accession IDs of the sequences included in the molecular phylogenetic trees are listed in [supplementary table S2](#), [Supplementary Material](#) online.

## Abstract

We have recently shown that the human *Nuclear pore-associated protein (NPAP1)/C15orf2* gene encodes a nuclear pore-associated protein. This gene is one of several paternally expressed imprinted genes in the genomic region 15q11q13. Because the Prader–Willi syndrome is known to be caused by the loss of function of paternally expressed genes in 15q11q13, a phenotypic contribution of *NPAP1* cannot be excluded. *NPAP1* appears to be under strong positive Darwinian selection in primates, suggesting an important function in primate biology. Interestingly, however, in contrast to all other protein-coding genes in 15q11q13, *NPAP1* has no ortholog in the mouse. Our investigation of the evolutionary origin of *NPAP1* showed that the gene is specific to primate species and absent from the 15q11q13-orthologous regions in all nonprimate mammals. However, we identified a group of paralogous genes, which we call *NPAP1L*, in all placental mammals except rodents. Phylogenetic analysis revealed that *NPAP1*, *NPAP1L*, and another group of genes (*UPF0607*), which is also restricted to primates, are closely related to the vertebrate transmembrane nucleoporin gene *POM121*, although they lack the transmembrane domain. These three newly identified groups of genes all lack conserved introns, and hence, are likely retrogenes. We hypothesize that, in the common ancestor of placentals, the *POM121* gene retrotransposed and gave rise to an *NPAP1*-ancestral retrogene *NPAP1L/NPAP1/UPF0607*. Our results suggest that the nuclear pore-associated gene *NPAP1* originates from the vertebrate nucleoporin gene *POM121* and—after several steps of retrotransposition and duplication—has been subjected to genomic imprinting and positive selection after integration into the imprinted *SNRPN-UBE3A* chromosomal domain.

**Key words:** C15orf2, NPAP1L, UPF0607, imprinting, nucleoporin, nuclear pore complex.

## Background

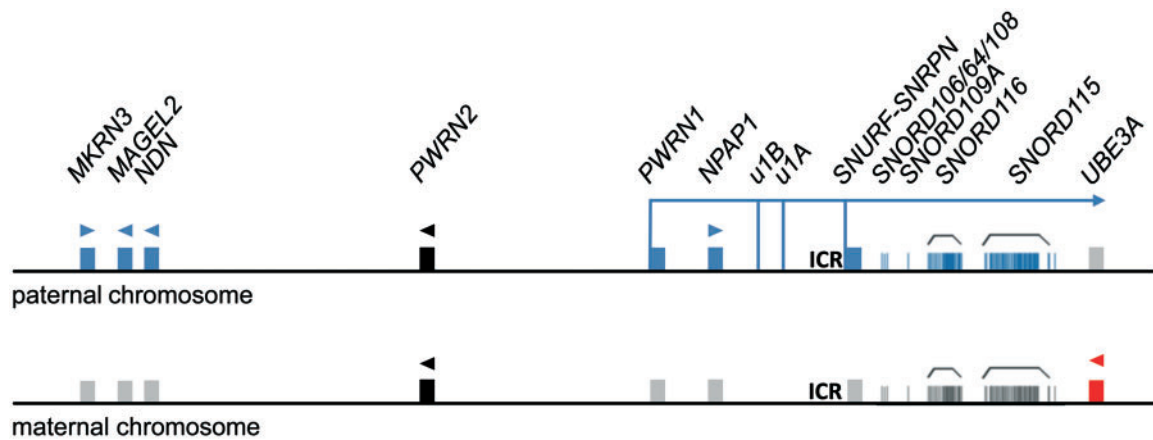
Genomic imprinting is a process that regulates the expression of certain genes in a parent-of-origin-dependent manner and has evolved independently in plants and mammals (Feil and Berger 2007). Genes regulated by this epigenetic mechanism are thus expressed either from the maternal chromosome or the paternal chromosome only. Most imprinted genes are organized in clusters, which are regulated by imprinting control regions (ICRs) that regulate the monoallelic expression of the genes in cis (Reik and Maher 1997; Ferguson-Smith 2011).

The human genomic region 15q11q13 contains a cluster of imprinted genes with several paternally only expressed genes

and one maternally only expressed gene (fig. 1). It is regulated by an ICR that includes the promoter of the *SNRPN* gene (Buiting et al. 1995; Saitoh et al. 1996; Ohta et al. 1999; Horsthemke and Buiting 2008). Loss of function of the paternally expressed genes in this region, most commonly arising through a ~6 Mb deletion on the paternally inherited chromosome, leads to Prader–Willi syndrome with neonatal muscular hypotonia and failure to thrive, childhood-onward hyperphagia and obesity, and mild-to-moderate intellectual disability (Cassidy 1997; Butler and Palmer 1983; Buiting 2010). We have recently shown that the paternally expressed gene *C15orf2* encodes a nuclear pore complex (NPC) associated protein, and it was therefore renamed to *nuclear*

© The Author(s) 2014. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)



**Fig. 1.**—The imprinted gene cluster in the human genomic region 15q11q13. The region, shown from centromere to telomere, contains a number of genes expressed from the paternal chromosome only (blue) and the maternally only expressed gene *UBE3A* (red). The direction of transcription is indicated by arrowheads. Nonexpressed alleles are depicted in gray on the repressed allele. Bially expressed genes are depicted in black. The complex *SNURF/SNRPN* locus has several alternative transcription start sites and encodes two proteins (SNURF and SNRPN), several snoRNAs, and a *UBE3A*-antisense RNA. The existence of a continuous *SNRPN* transcript (blue arrow) containing upstream and downstream parts has, however, not yet been experimentally documented. The figure shows the gene expression according to fetal brain and is only approximately drawn to scale.

*pore-associated protein 1* (*NPAP1*) (Neumann et al. 2012). The exact function of the protein is not known, but we suspect a brain-specific NPC-related function because the protein is expressed in several human brain regions (Wawrzik et al. 2010). Several studies showed that *NPAP1* underwent strong positive Darwinian selection in the primate lineage (Nielsen et al. 2005; Kosiol et al. 2008; Wawrzik et al. 2010).

The imprinted domain in 15q11q13 has assembled relatively recently during mammalian evolution from before unlinked and nonimprinted components (Rapkins et al. 2006). The region evolved its imprinted regulation after fusion of two nonimprinted regions that contained *SNRPN* and *UBE3A*, respectively, 105–180 Ma (Rapkins et al. 2006). Other intronless genes, such as *MKRN3*, *MAGEL2*, and *NDN* integrated independently into this genomic region after fusion of *SNRPN* and *UBE3A*, most probably by retrotransposition (Gray et al. 2000; Chai et al. 2001; Rapkins et al. 2006). It was hypothesized that at least some of the retrogenes integrated into the region after the evolution of imprinting in 15q11q13 and acquired their imprinted regulation subsequently (Chai et al. 2001; Rapkins et al. 2006). In the well-described murine orthologous region, two rodent-specific imprinted genes, *Frat3/Peg12* and *Atp5l-ps1*, have been identified, suggesting that the process of gene acquisition is still ongoing and leads to divergent imprinted gene sets in the primate and rodent lineage (Chai et al. 2001). Here, we report that *NPAP1* is a primate-specific gene that entered the imprinted region 15q11q13 by duplication from an ancestral paralog on human chromosome 9 during primate evolution. The ancestral gene, *NPAP1L*, in turn is derived from retrotransposition of *POM121* in an ancestor of placentals.

## Materials and Methods

### Bioinformatic Tools

*NPAP1* homologous gene sequences were found using the “orthologs” list in the Ensembl database ([www.ensembl.org](http://www.ensembl.org), last accessed January 20, 2014) or using the Blast-like alignment tool (BLAT) in the University of California, Santa Cruz (UCSC) genome browser (<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>, last accessed January 20, 2014). Gene and protein alignments were produced using the ClustalW multiple alignment tool included in the Geneious Pro 5.6 package (Biomatters Ltd, Auckland, New Zealand) with standard settings. The prediction of intronless ORFs was also carried out with Geneious Pro 5.6 or the ORF finder in National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>, last accessed January 20, 2014). Exon prediction was performed with the software GENSCAN (<http://genes.mit.edu/GENSCAN.html>, last accessed January 20, 2014) and standard settings.

### Sequence Analyses

Marmoset *NPAP1* sequencing was performed with Big Dye Terminators (BigDye Terminator v1.1 Cycle Sequencing Kit, Life Technologies, Darmstadt, Germany) and the cycle sequencing procedure. Products were analyzed with an ABI 3100 Genetic Analyzer and Sequencing Analysis software (Life Technologies). The following primers were used: Marmoset\_NPAP1\_fw1: AAACACCCCAGCTCCGTGAGGA; Marmoset\_NPAP1\_rev1: GGATGGGCTGGGAAGTTGTGGC; Marmoset\_NPAP1\_fw2: CACAACAGGCCCTGCAAAAGGA; Marmoset\_NPAP1\_rev2: CCCCATGTAAAACGGGAGGCAC;

Marmoset\_NPAP1\_fw3: ATCCAATTCTGGGGCTCTTG;  
Marmoset\_NPAP1\_rev4: TCCAAGGTGCCAGGTCTC.

### Expression Analyses

For expression analysis of *NPAP1L*, we used a customized panel of cDNAs from bovine tissues (caudate nucleus, cerebellum, cerebral cortex, hippocampus, hypothalamus, kidney, liver, placenta, skeletal muscle, testis) and DNA from bovine testis as a control (Zyagen, San Diego, USA). Polymerase chain reaction (PCR) assays were designed such that they spanned introns from the Ensembl gene prediction. For intron 1 (1,067 bp), we used the primers Cow-NPAP1L-in1-fw2: TAACTATCCCTTTGACTCCCGA and Cow-NPAP1L-in1-rev2: CTG GAGCATAGATAACTGCCAA. For intron 2 (17 bp), we used the primers Cow-NPAP1L-in2-fw: CAAGCCTCAACTATTTCCTTG and Cow-NPAP1L-in2-rev: TGGCAAACCTGAATCCATTTTG. Representative products were sequenced using Applied Biosystems BigDye Terminator v1.1 Cycle Sequencing Kit (Life Technologies) and an ABI PRISM 3100 Genetic Analyzer (Life Technologies). The control PCR on bovine *ATCB* was performed with primers Fw2-beta-Actin-Cow: GGCACCCAGCACAATGAAGA and Rev2-beta-Actin-Cow: CGACTGCTGTACCTTCACCG.

### Phylogenetic Analyses

BlastP searches at NCBI (<http://blast.ncbi.nlm.nih.gov>, last accessed January 20, 2014) were performed using the peptide sequence of NPAP1 as query. The ML tree based on the JTT +  $\Gamma_4$  amino acid substitution model was calculated using the MEGA5 program (Tamura et al. 2011) and the embedded alignment software MUSCLE (Edgar 2004), with a data set including all obtained sequences plus homologs identified by Ensembl.

## Results and Discussion

### Conservation of *NPAP1* Orthologs in Primates

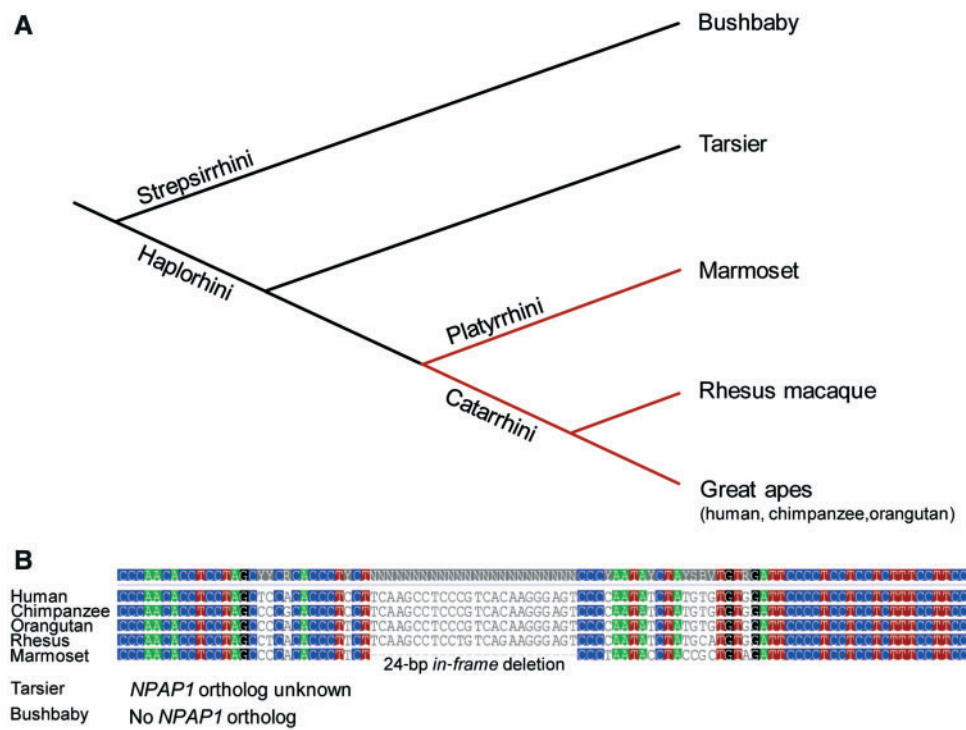
Ensembl (release 69; <http://www.ensembl.org>, last accessed January 20, 2014; [Hubbard et al. 2009]) contains *NPAP1* orthologs in the chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), orangutan (*Pongo abelii*), and rhesus macaque (*Macaca mulatta*). In an effort to identify additional *NPAP1* orthologs in the 15q11q13 orthologous regions of other primate species, we searched the UCSC genome browser (<http://genome.ucsc.edu>, last accessed January 20, 2014; [Kent 2002]) using the BLAT Kent et al. [2002]) with the human gene sequence as query. This search resulted in the identification of *NPAP1* orthologs in the gibbon (*Nomascus leucogenys*), the squirrel monkey (*Saimiri boliviensis*), and the marmoset (*Callithrix jacchus*). But in the genome of the bushbaby (*Otolemur garnettii*), a member of the more distantly related suborder of Strepsirrhini (fig. 2), we could not find an *NPAP1*-homologous sequence in the 15q11q13-

orthologous region. The genomes of baboon (*Papio anubis*), tarsier (*Tarsius syrichta*), and mouse lemur (*Microcebus murinus*) are not sufficiently assembled to determine syntenic relationships.

Of all analyzed primates, the marmoset and the squirrel monkey, members of the Haplorhini parvorder Platyrrhini, were the most distantly related species that already have orthologs of *NPAP1* (fig. 2). An alignment showed that the marmoset *NPAP1* gene on chromosome 6 contains a ~2.5 kb gap in the part that aligns with the human open reading frame (ORF). We thus sequenced the missing part of the gene in the marmoset and obtained a complete gene sequence (supplementary table S1, Supplementary Material online), that is, 69.9% identical with the human *NPAP1*. In contrast to *NPAP1* orthologs of Catarrhini species, the marmoset *NPAP1* lacks a long intronless ORF. GENSCAN (<http://genes.mit.edu/GENSCAN.html>, last accessed January 20, 2014, [Burge and Karlin 1997]) predicts a 3.06 kb ORF with two small introns that starts 164 bp upstream of the human ORF and ends parallel at an apparently homologous position (supplementary table S1, Supplementary Material online). The second and third exon of the GENSCAN prediction contain six deletions when compared with the Catarrhini genomes that are multiples of 3 bp, and up to 39 bp long, so would appear to be in frame and coding. By contrast, the upstream part of the gene contains a number of small indels that are not multiples of three, suggesting that this region is not part of the ORF and might have become partially pseudogenized. The indel pattern suggests a balancing selection against frame-shifts inside the coding sequence, as 1 bp indels occur with approximately ten times higher probability than 3 bp indels (de la Chaux et al. 2007). This argues for a protein coding function of marmoset *NPAP1* despite the absence of the expected intronless ORF. As *NPAP1* orthologs were found in all analyzed members of Haplorhini but not in any members of Strepsirrhini (fig. 2), the gene presumably integrated into the 15q11q13 orthologous region after the two primate suborders diverged about 60–70 Ma (Springer et al. 2012).

### A New Family of *NPAP1* Homologous Sequences in Placental Mammals

In addition to the aforementioned primate *NPAP1* genes, Ensembl (release 69) also contains apparently homologous genes from dog (ENSCAFG00000014649 and ENSCAFG0000 0023307), cow (ENSBTAG00000046462), pig (ENSSSCG00000022532), and elephant (ENSLAFG00000014287 and ENSLAFG00000031777) that were annotated as *NPAP1* orthologs. In light of the known absence of *NPAP1* genes from the murine orthologous region, this annotation and phylogenetic distribution seemed to contradict the traditional view of mammalian evolution, which places rodents closer to humans than ruminants and carnivores (e.g., Springer et al. 2003). Alternatively, but less



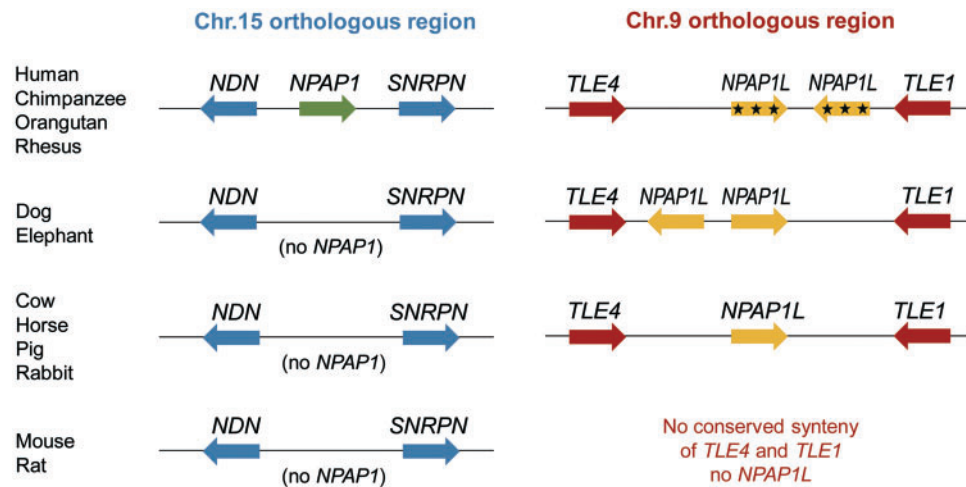
**Fig. 2.**—*NPAP1* conservation in primates. (A) Cladogram showing the family structure of primates and the relationship of the analyzed primate species. *NPAP1* orthologs were found in all analyzed members of the parvorders Platyrrhini and Catarrhini (red). (B) Detailed view of an alignment of selected primate *NPAP1* orthologs. The figure shows a well-conserved region inside the human ORF (1,965–2,062 bp from NCBI reference sequence NM\_018958.2) that contains one of the numerous *in-frame* deletions found in the marmoset sequence.

parsimoniously, the nonprimate homologs might be of different origin. Therefore, we investigated the conservation of synteny among the Ensembl orthologs more closely. Using the UCSC genome browser, we found that only primate *NPAP1* genes share synteny with *UBE3A* and *SNRPN*, whereas the homologous genes in other mammals are located between their respective orthologs of *transducin-like enhancer of split (TLE) 1* and *TLE4* (fig. 3). We refer to the group of orthologous genes located in this synteny group as *nuclear pore-associated protein 1-like (NPAP1L)*. *TLE1* and *4* belong to a larger region of well-conserved synteny including orthologs of the human genes *GNA14*, *GNAQ*, *PSAT1*, *TLE4*, *TLE1*, *RASEF*, and *FRMD3* (centromeric to telomeric on human chr9q21). Because the complete synteny group can also be found in chicken, which does not contain *NPAP1L*, it is likely that *NPAP1L* integrated into this region as a single gene during mammalian evolution. A possible mechanism for this would be retrotransposition (Kaessmann 2010).

Ensembl (release 69) also contains genes from three additional mammalian species (cat, ENSFCAG00000010320; tree shrew, ENSTBEG00000007225; ferret, ENSMPUG00000019448) that were annotated as *NPAP1*. The current quality of the genome assemblies of these species does not allow us to conduct robust synteny analyses. However, cat scaffold GL897178.1 contains both *TLE1* and ENSFCAG0000

0010320, suggesting that ENSFCAG00000010320 is another ortholog of *NPAP1L*. All Ensembl-annotated *NPAP1L* orthologs were predicted to be protein coding by the Ensembl pipeline (Potter et al. 2004).

Using the UCSC genome browser and the BLAT algorithm with human *NPAP1* as query, we found *NPAP1L* sequences in the genomes of additional mammals, for example, horse and rabbit. We could, however, not identify significant hits in either rodents or marsupials. In the genomes of mouse and rat, *TLE4* (on murine chr.19) and *TLE1* (on murine chr. 4) have lost their syntenic position, creating the impression that *NPAP1L* might have been lost in their common ancestor during chromosomal rearrangements. In the guinea pig genome, however, the *TLE4-TLE1* synteny group is found on scaffold 21 although *NPAP1L* was not identified, rather suggesting that *NPAP1L* has been lost in rodents before chromosomal rearrangements in this region. In the human genome, our search revealed two copies of *NPAP1L* in opposite orientations on human chromosome 9 that match Ensembl-annotated processed pseudogenes (ENSG00000238002 on the plus strand and ENSG00000236521 on the minus strand). The same arrangement of *NPAP1L* genes was also observed in other primate species (fig. 3). Both human *NPAP1L* copies contain only short ORFs, which code for peptides without significant amino acid sequence similarity to



**Fig. 3.**—Synteny conservation in mammals. *NPAP1* and *NPAP1L* orthologs lie in two different synteny groups, which are orthologous to human 15q11q13 or 9q21, respectively. The arrows show the orientation of genes from 5' to 3'. Mutations leading to pseudogenization of primate *NPAP1L* genes are represented by stars. Because the relative orientations of the two neighboring *NPAP1L* copies differ between primates and dog/elephant, these duplications do not appear to have a common origin.

*NPAP1* and *NPAP1L* proteins, confirming their annotation as processed pseudogenes. As is the case for the human genome, also the genomes of dog and elephant each contain two highly similar copies of *NPAP1L* with opposite orientations. However, in contrast to dog and elephant, the human *NPAP1L* copies are tail-to-tail oriented, suggesting that *NPAP1L* tandem-duplicated several times independently (fig. 3).

#### Expression Analysis of *NPAP1L*

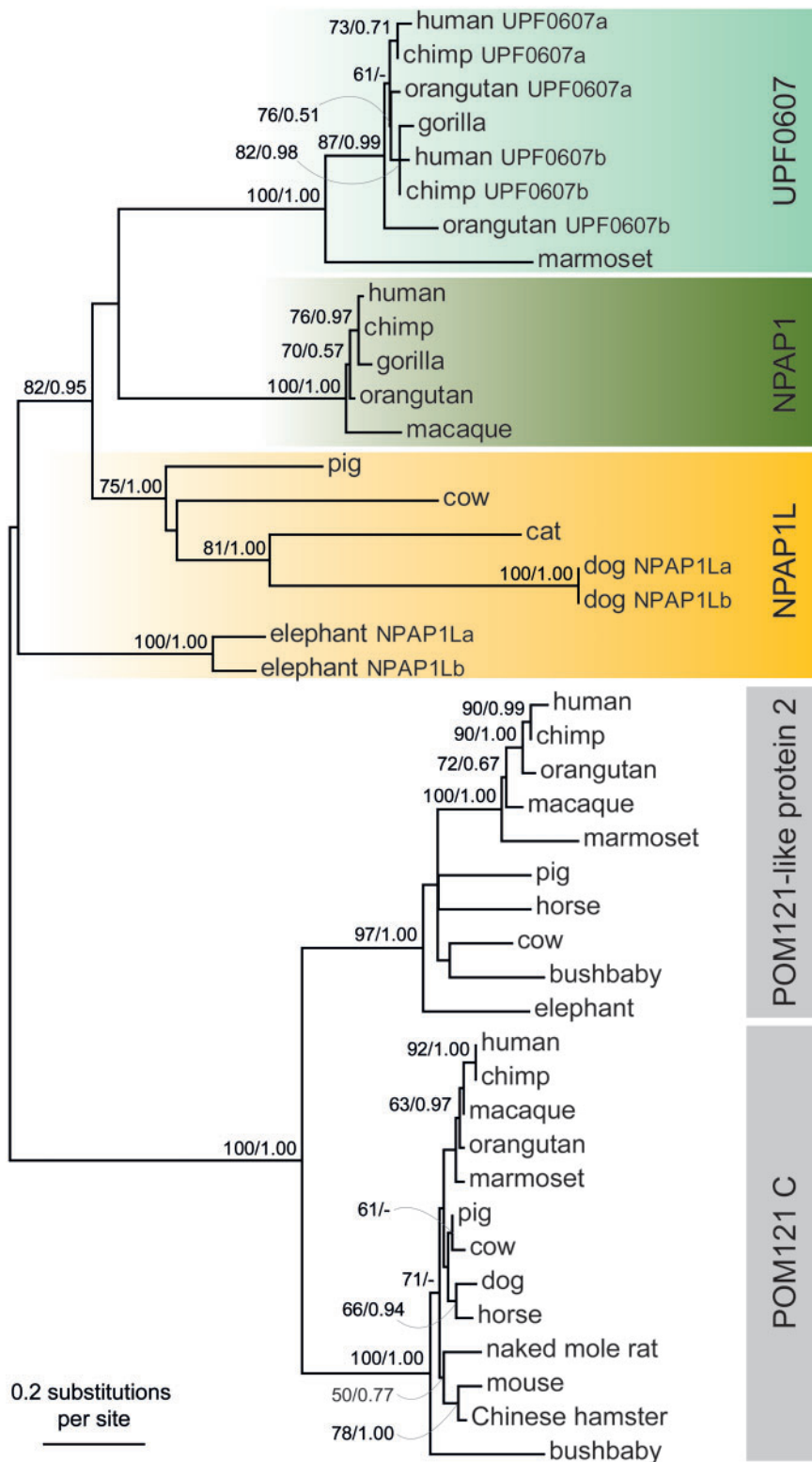
As mentioned earlier, all nonprimate *NPAP1L* orthologs were predicted to be protein coding by Ensembl. In an effort to verify the in silico prediction in a nonprimate species, we analyzed the expression of bovine *NPAP1L*. To this end, we used a bovine cDNA panel (Zyagen) and two PCR assays spanning the two introns of the Ensembl-predicted gene (ENSBTAG00000046462). As we had shown before that human *NPAP1* is expressed in testis and brain (Färber et al. 2000; Wawrzik et al. 2010), we included cDNA from different bovine brain regions and testis into our panel. With both PCR assays, we observed expression of bovine *NPAP1L* in four of five analyzed brain regions: caudate nucleus, cerebellum, hippocampus, and hypothalamus. With one of the two assays, we also obtained weak signals for the cerebral cortex, kidney, and testis, but we did not observe expression in liver, placenta, or skeletal muscle (data not shown). Curiously, by gel analysis and Sanger sequencing, the obtained products from cDNA did not correspond to mRNA of the expected splicing pattern but were colinear with genomic DNA.

Although our cDNA panel had been obtained from a commercial source (Zyagen) that tests for residual DNA contamination as part of their quality control procedure, the

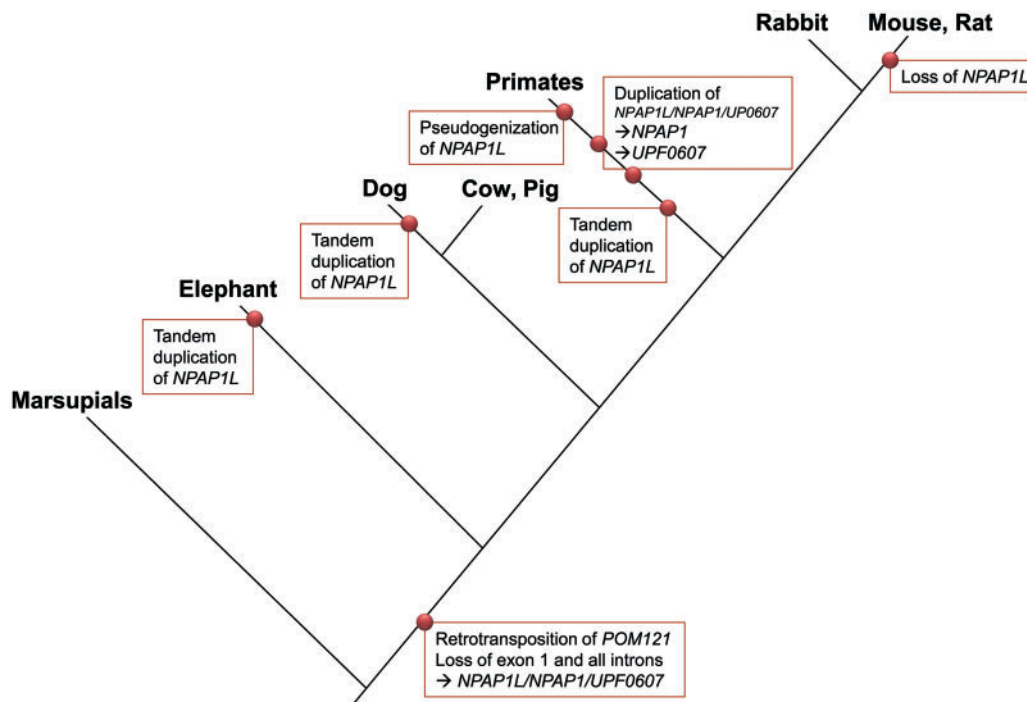
unexpected finding of colinear RNA expression prompted us to double-check for contamination with genomic DNA. To this end, we used an intron-spanning RT-PCR for the *ACTB* locus that gives rise to a 131-bp larger product when genomic DNA is amplified. In all cDNA samples, we only obtained the product that is expected from *ACTB* cDNA, making a contamination of the cDNAs with genomic DNA very unlikely. We conclude that *NPAP1L* is expressed in the cow, primarily in the brain, but that the splicing pattern predicted by Ensembl is not correct in any of the analyzed tissues. Alternatively, it is possible that a 2,688-bp intronless ORF is expressed in the cow and would lead to a shorter *NPAP1*-homologous protein of 895 amino acids. Even if the bovine *NPAP1L* ortholog should not be protein coding, this would not per se exclude protein-coding functions for other mammalian *NPAP1L* orthologs.

#### *NPAP1* and *NPAP1L* Belong to a *POM121*-Related Gene Family

Because functional *NPAP1L* sequences could be found in all analyzed mammals except rodents, whereas *NPAP1* is unique to primate species, we hypothesized that *NPAP1* originates from the *NPAP1L* gene locus. In order to investigate this hypothesis in an unbiased way, we performed a BlastP (<http://blast.ncbi.nlm.nih.gov>, last accessed January 20, 2014) search with the human *NPAP1* protein sequence as query. Sequences obtained in this search as well as Ensembl-annotated protein sequences were used for the inference of phylogenetic relationships. A maximum likelihood (ML) tree (fig. 4) was reconstructed using representatives of major tetrapod taxa. This data set consisted of *NPAP1* amino acid sequences from different primate species, of a group of proteins designated as UPF0607 that are also restricted to primates, of *NPAP1L*



**Fig. 4.**—ML tree of NPAP1-homologous proteins. Boot strap values (left) and associated support values (right) are given for all nodes that have bootstrap values above 50. The tree topology suggests that *NPAP1*, *UPF0607*, and *NPAP1L* are derived from the vertebrate-specific nucleoporin gene *POM121*. *NPAP1L* seems to be ancestral to the primate-specific sister genes *NPAP1* and *UPF0607*. On the basis of branch lengths, it can be seen that the two neighboring *NPAP1L* copies (*NPAP1La* and *NPAP1Lb*) in dog and elephant are the result of two independent duplications.



**Fig. 5.**—Cladogram showing the timing of evolutionary events during the formation of the *NPAP1*-related gene family. The red dots mark the approximate phylogenetic timings of gene duplications and losses that must be minimally assumed to explain the gene distribution in the mammalian family. In the primate branch, the marked four events must have taken place although, based on our results, their exact order cannot be defined.

proteins from elephant, dog, cat, pig, and cow, and of distantly related *POM121* and *POM121*-like proteins (supplementary table S2, Supplementary Material online).

The inferred ML tree suggests that the genes *NPAP1L*, *NPAP1*, and *UPF0607* form a monophyletic group that is derived from *POM121* (fig. 4). The most straightforward interpretation of the tree topology and the taxonomic distribution of genes is that *NPAP1L* duplicated in the lineage leading to primates, resulting in an *NPAP1/UPF0607* gene. Before the primate radiation, this ancestral *NPAP1/UPF0607* gene duplicated again to give rise to *NPAP1* and *UPF0607*. The primate-specific *UPF0607* gene duplicated in a subset of primate species resulting in two groups of genes, namely *UPF0607a* and *-b* (see fig. 4 and supplementary table S2, Supplementary Material online). In dog and elephant, two copies of *NPAP1L* genes derived from recent independent duplication events are found, namely *NPAP1La* and *-b* (see fig. 4 and supplementary table S2, Supplementary Material online). In the lineage leading to primates, the ancestral *NPAP1L* gene pseudogenized and is therefore not included in our phylogenetic analysis. However, remnants of this gene are still identifiable in all analyzed primate species. Taken together, the tree topology supports our initial hypothesis that *NPAP1* is derived from the *NPAP1L* gene locus and adds *POM121* as a common ancestor of both groups of genes.

Although the vertebrate nucleoporin gene *POM121* contains several large and highly conserved introns, all *NPAP1* and

*UPF0607* genes are intronless. *NPAP1L* genes are predicted to contain small introns; however, the intronic structure is not conserved between the orthologs and presumably evolved secondarily from an intronless ancestral gene. In light of these observations, the phylogenetic analysis suggests an evolutionary scenario wherein *POM121* duplicated via retrotransposition in the last common ancestor of placentals, giving rise to the intronless *NPAP1L/NPAP1/UPF0607* retrogene (fig. 5). The first exon of *POM121* that includes its transmembrane domain was lost during or following retrotransposition, because none of the orthologs of *NPAP1L*, *NPAP1*, or *UPF0607* are predicted transmembrane proteins. The loss of 5' gene parts is characteristic of gene duplications via retrotransposition (Ding et al. 2006). More recently, before the primate radiation, this gene duplicated or retrotransposed twice more giving rise to the genes *NPAP1* and *UPF0607*. In dog and elephant, *NPAP1L* was subject to independent tandem duplications, whereas it was lost in the rodent lineage (fig. 5).

## Conclusions

Our results show that the imprinted, primate-specific *NPAP1* gene originates from the vertebrate nucleoporin gene *POM121*. It is part of a so far unrecognized gene family of *POM121*-related retrogenes, the members of which can be considered possible candidates for mammal- and

primate-specific NPC-associated functions. Unlike POM121, the predicted proteins lack a transmembrane domain and thus appear to have functionally diverged from their ancestral protein.

## Supplementary Material

Supplementary tables S1 and S2 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

The authors thank Dr Jörg Gromoll for providing marmoset DNA. This work was supported by the Deutsche Forschungsgemeinschaft (grant number HO 949/19-1).

## Literature Cited

- Buiting K. 2010. Prader-Willi syndrome and Angelman syndrome. *Am J Med Genet. C* 154:365–376.
- Buiting K, et al. 1995. Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nat Genet.* 9:395–400.
- Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *J Mol Biol.* 268:78–94.
- Butler MG, Palmer CG. 1983. Parental origin of chromosome 15 deletion in Prader-Willi syndrome. *Lancet* 1:1285–1286.
- Cassidy SB. 1997. Prader-Willi syndrome. *J Med Genet.* 34:917–923.
- Chai J-H, Locke DP, Ohta T, Grealley JM, Nicholls RD. 2001. Retrotransposed genes such as *Frat3* in the mouse Chromosome 7C Prader-Willi syndrome region acquire the imprinted status of their insertion site. *Mamm Genome.* 12:813–821.
- de la Chaux N, Messer PW, Arndt PF. 2007. DNA indels in coding regions reveal selective constraints on protein evolution in the human lineage. *BMC Evol Biol.* 7:191.
- Ding W, Lin L, Chen B, Dai J. 2006. L1 elements, processed pseudogenes and retrogenes in mammalian genomes. *IUBMB Life* 58:677–685.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Färber C, Gross S, Neesen J, Buiting K, Horsthemke B. 2000. Identification of a testis-specific gene (*C15orf2*) in the Prader-Willi syndrome region on chromosome 15. *Genomics* 65:174–183.
- Feil R, Berger F. 2007. Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet.* 23:192–199.
- Ferguson-Smith AC. 2011. Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet.* 12:565–575.
- Gray TA, et al. 2000. The ancient source of a distinct gene family encoding proteins featuring RING and C(3)H zinc-finger motifs with abundant expression in developing brain and nervous system. *Genomics* 66:76–86.
- Horsthemke B, Buiting K. 2008. Genomic imprinting and imprinting defects in humans. *Adv Genet.* 61:225–246.
- Hubbard TJ, et al. 2009. Ensembl 2009. *Nucleic Acids Res.* 37:D690–D697.
- Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* 20:1313–1326.
- Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res.* 12:656–664.
- Kent WJ, et al. 2002. The human genome browser at UCSC. *Genome Res.* 12:996–1006.
- Kosiol C, et al. 2008. Patterns of positive selection in six mammalian genomes. *PLoS Genet.* 4:e1000144.
- Neumann LC, et al. 2012. The imprinted NPAP1/C15orf2 gene in the Prader-Willi syndrome region encodes a nuclear pore complex associated protein. *Hum Mol Genet.* 21:4038–4048.
- Nielsen R, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 3:e170.
- Ohta T, et al. 1999. Imprinting-mutation mechanisms in Prader-Willi syndrome. *Am J Hum Genet.* 64:397–413.
- Potter SC, et al. 2004. The Ensembl analysis pipeline. *Genome Res.* 14:934–941.
- Rapkins RW, et al. 2006. Recent assembly of an imprinted domain from non-imprinted components. *PLoS Genet.* 2:e182.
- Reik W, Maher ER. 1997. Imprinting in clusters: lessons from Beckwith-Wiedemann syndrome. *Trends Genet.* 13:330–334.
- Saitoh S, et al. 1996. Minimal definition of the imprinting center and fixation of chromosome 15q11-q13 epigenotype by imprinting mutations. *Proc Natl Acad Sci U S A.* 93:7811–7815.
- Springer MS, et al. 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLoS One* 7:e49521.
- Springer MS, Murphy WJ, Eizirik E, O'Brien SJ. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc Natl Acad Sci U S A.* 100:1056–1061.
- Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.
- Wawrzik M, et al. 2010. The *C15orf2* gene in the Prader-Willi syndrome region is subject to genomic imprinting and positive selection. *Neurogenetics* 11:153–161.

Associate editor: B. Venkatesh