



## Linking dopamine neurotransmission and neurogenesis: The evolutionary history of the NTAD (*NCAM1-TTC12-ANKK1-DRD2*) gene cluster

Nina Roth Mota\*, Eli Vieira Araujo-Jnr\*, Vanessa Rodrigues Paixão-Côrtes, Maria Cátira Bortolini and Claiton Henrique Dotto Bau

*Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.*

### Abstract

Genetic studies have long suggested the important role of the *DRD2* gene in psychiatric disorders and behavior. Further research has shown a conjoined effect of genes in the Chr11q22-23 region, which includes the *NCAM1*, *TTC12*, *ANKK1* and *DRD2* genes, or NTAD cluster. Despite a growing need to unravel the role of this cluster, few studies have taken into account interspecies and evolutionary approaches. This study shows that behaviorally relevant SNPs from the NTAD cluster, such as rs1800497 (Taq1A) and rs6277, are ancient polymorphisms that date back to the common ancestor between modern humans and Neanderthals/Denisovans. Conserved synteny and neighborhood indicate the NTAD cluster seems to have been established at least 400 million years ago, when the first Sarcopterygians emerged. The NTAD genes are apparently co-regulated and this could be attributed to adaptive functional properties, including those that emerged when the central nervous system became more complex. Finally, our findings indicate that NTAD genes, which are related to neurogenesis and dopaminergic neurotransmission, should be approached as a unit in behavioral and psychiatric genetic studies.

*Keywords:* NTAD cluster, Taq1A SNP, shared synteny, co-regulation, psychiatric genetics.

### Introduction

#### The role of the NTAD cluster in psychiatric disorders

The crucial role played by dopaminergic neurotransmission systems in psychiatric genetics has been of great interest for several years. Among the candidate genes, one of the main focuses of research has been the dopamine receptor D2 (*DRD2*), especially its neighboring single-nucleotide polymorphism (SNP) rs1800497 (Taq1A). The rs1800497 SNP was considered to be a silent mutation located about 10 kb from *DRD2*, in the 3' untranslated region of this gene. The identification of a novel gene in the neighboring forward-strand region of *DRD2*, named *ANKK1*, showed that the rs1800497 SNP is actually located in exon 8 of *ANKK1* (Neville *et al.*, 2004), where it causes an amino acid change (Glu713Lys) in its 11<sup>th</sup> ankyrin repeat. Although the rs1800497 polymorphism is localized in *ANKK1*, it seems to be in linkage disequilibrium with several *DRD2* variants (Gelernter *et al.*, 2006; Dubertret *et al.*, 2010).

Despite the enormous number of studies regarding the role of the rs1800497 SNP in psychiatric disorders (es-

pecially alcohol and nicotine dependence), it is still far from clear how and to what degree it affects psychiatric disorders. The three most recent meta-analyses support the link between the rs1800497 T allele and alcoholism (Munafò *et al.*, 2007; Smith *et al.*, 2008; Le Foll *et al.*, 2009). There are two meta-analysis studies focusing on the relationship between rs1800497 and smoking behavior, but with conflicting results (Li *et al.*, 2004; Munafò *et al.*, 2004). Strong heterogeneity has been a hallmark of several of these meta-analyses. Consequently, the strong divergence among such findings raises the need to identify the variables that might explain such heterogeneity.

It has been suggested that two other nearby genes, *NCAM1* and *TTC12*, are also good candidate genes for psychiatric disorders and could comprise causative variants to phenotypes previously attributed to *DRD2* polymorphisms (Gelernter *et al.*, 2006; Huang *et al.*, 2008; Dubertret *et al.*, 2010). For example, Yang *et al.* (2007) found an association of SNPs in the *NCAM1*, *TTC12* and *ANKK1* genes with alcohol dependence, but not so for SNPs in the *DRD2* gene. In another study, a single haplotype spanning *TTC12* and *ANKK1*, as well as multiple SNPs in these two genes, were associated with nicotine dependence (Gelernter *et al.*, 2006).

These genes are located on chromosome 11 (more precisely, the 11q22-23 region) and form a 521 kb gene

Send correspondence Claiton H.D. Bau. Departamento de Genética, Instituto de Biociências, UFRGS, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. E-mail: claiton.bau@ufrgs.br.

\*These authors contributed equally to the present study.

cluster that comprises the *NCAMI*, *TTC12*, *ANKK1* and *DRD2* genes, known as the NTAD gene cluster (Figure S1). The four genes that comprise the NTAD cluster all seem to act on the brain, although details of their specific functions in neural tissue have yet to be discovered. The neural cell adhesion molecule 1 (*NCAMI*) plays an important role in neurogenesis, specifically in axon and dendrite growth (McIntyre *et al.*, 2010). *TTC12* encodes the poorly understood tetratricopeptide repeat domain 12 protein, which seems to be involved in dopaminergic transmission and neurodevelopment via the Wnt signaling pathway (Kahto and Kahto, 2003; Castelo-Branco and Arenas, 2006). The ankyrin repeat and kinase domain containing 1 (*ANKK1*) gene encodes a signaling protein which takes part in indirect modulation of the expression of *DRD2* (Huang *et al.*, 2008), thus constituting the clearest currently known evidence of co-regulation in the NTAD cluster.

### Gene cluster and genomic architecture: Evolutionary aspects

Gene order in eukaryotes cannot be attributed in its entirety to mere randomness. Multiple lines of evidence indicate that co-expressed, co-regulated and co-functional genes can be maintained as a gene cluster due to the pressure of natural selection (Hurst *et al.*, 2004; Sémon and Duret, 2006; Michalak, 2008).

Considering that polymorphisms of the NTAD gene cluster seem to have a conjoined effect, the question raised is how long these genes are being maintained with shared synteny (genes in the same chromosome) and conserved neighborhood (genes side-by-side in the same order) through the course of evolution. If the NTAD conformation is recent or human-specific it may reflect an adaptive novelty or stochastic clustering in *Homo/Homo sapiens*. Notwithstanding, the maintenance of a specific gene cluster such as NTAD over long evolutionary periods is much more difficult to be explained by actions of random processes only. An ancient clustering in this case might also reflect a functional benefit and/or an orchestrated evolution of genes involved in neurotransmission and neurogenesis, probably mediated by co-regulation, co-expression or molecular co-functionality.

Hominid comparative genomics has made great strides after the publication of the Neanderthal and Denisovan nuclear genomes since it has allowed researchers to build on knowledge about unique human phenotypes, including psychiatric disorder susceptibilities in present-day populations. Some of these psychiatric disorders have been repeatedly associated with variation in the whole NTAD cluster, but no previous study investigated it from a wider evolutionary perspective using comparative genomic approaches. Taking this information into account we herein performed the first *in silico* study addressing the genomic architecture and chromosomal dynamics of the NTAD cluster from an evolutionary perspective. Our results pro-

vide a more comprehensive view about this gene cluster and how its dynamics could shape future genetic studies of complex behavior phenotypes and psychiatric disorder in humans.

## Material and Methods

### Analysis of NTAD SNP status in primates

Seven human SNPs were chosen based on previous association with psychiatric disorders and/or with evidence of functionality: rs646558 from *NCAMI*; rs723077 and rs2303380 from *TTC12*; rs2734849 and rs1800497 from *ANKK1*; and rs6277 and rs2283265 from *DRD2* (Figure S1). These polymorphic sites were then compared with their counterparts in the genomes of two archaic hominids, *Homo neanderthalensis* (Green *et al.*, 2010) and a Denisovan specimen (Reich *et al.*, 2010), as well as nine non-human primates. The derived allele age was estimated for the rs1800497 and rs6277 SNPs according to the frequency based method, as proposed by Slatkin and Rannala (2000).

Three different approaches were used to predict whether the three non-synonymous mutations (rs1800497 and rs2734849 in *ANKK1*; rs723077 in *TTC12*) among the seven SNPs promote important functional changes in the proteins: PolyPhen2, SNAP and the assessment of Grantham scores of chemical distance (Grantham, 1974). For the latter we used the classification by Li *et al.* (1985) as: conservative (Grantham score = 0-50), moderately conservative (51-100), moderately radical (101-150) and radical (> 151). These methods are used to predict the possible impact of amino acid substitutions on the structure and function of proteins by means of chemico-physical and comparative evidence.

### Orthology, synteny and neighborhood status of the NTAD cluster in vertebrate genomes

The online databases Ensembl release 66, UCSC Genome Browser and UniProt were used as sources for the nucleotide and protein sequences of *NCAMI*, *TTC12*, *ANKK1* and *DRD2* genes in the human genome and their orthologues in 47 other vertebrate species (Table S1). Their synteny and neighborhood status were inferred from the available contigs. BLAST/BLAT searches in these databases were performed to find possible unannotated orthologues. BioEdit version 7.0.9.0 (Hall, 1999) was used to align the sequences of the orthologues, when necessary. The DECODE database was used to infer transcription factor binding sites in the NTAD cluster. DECODE is based on text mining applications by SABiosciences and gene annotations of regulatory binding sites available at the UCSC Genome Browser.

## Results and Discussion

Comparative analyses of seven SNPs belonging to the NTAD gene cluster in human and non-human primates

Our analyses based on the seven SNPs selected according to their association with psychiatric disorders and/or with evidence of functionality, revealed interesting results (Table 1). For instance, we were able to show that polymorphisms believed to be *H. sapiens*-specific turned out to be plausibly widespread in the *Homo* genus. Although the introgression of *H. sapiens* derived alleles to other hominids could not be disregarded, these polymorphisms could be traced back to at least 270,000-800,000 years when our lineage diverged from Neanderthals/Denisovans (Green *et al.*, 2010; Reich *et al.*, 2010). This seems to be the case for the rs1800497 SNP, where both ancestral and derived alleles (taking into account that a mutation event produces a new, mutant or derived allele, that is different from the “original” or ancestral one) are present in the Denisovan genome, while in Neanderthals only the derived allele could be found, denoting that the A → G (Glu713Lys) mutation occurred before the origin of *Homo sapiens*. Our estimate for the age of the derived allele (Slatkin and Rannala, 2000) is compatible with this hypothesis (379,000-447,000 years).

A similar situation was detected with the *DRD2* rs6277 SNP, but in this case both alleles were found in Neanderthals (Table 1). Interestingly, the estimation of allele age for the derived allele of this polymorphism presented a greater discrepancy among populations. For Africans and Asians it was about 65,000 years, whereas for Europeans it was set at 359,000 years. These results raised some instigating hypotheses, including allele introgression from Neanderthals to *H. sapiens* in European populations, with subsequent dispersion to other continents. Selection processes and genetic drift, inflating the derived allele frequency in Europe, are other possible explanations. It seems however unlikely that stochastic processes would keep these polymorphisms unchanged for such a long time. Signals of positive selection were found in variant alleles of a related dopaminergic gene (*DRD4*) associated with modern psychiatric disorders. Some of these variants are associated with behavioral traits that could have had some adaptive advantage in the past, but today may have clinical implications (Ding *et al.*, 2002; Wang *et al.*, 2004; Tovo-Rodrigues *et al.*, 2011). A similar scenario may be responsible for the maintenance of these polymorphic sites in the human lineage, although only further studies can test this hypothesis.

On the other hand, the *TTC12* rs2303380 and *DRD2* rs2283265 SNPs are likely to be *Homo sapiens*-specific. The *NCAM1* rs646558 derived allele (C) is also *H. sapiens*-specific, but curiously its ancestral allele (A) is exclusively found in Old World primates, while a third allele (G) is found in lemurs and New World monkeys (Table 1). This

**Table 1** - Single nucleotide variation in the NTAD cluster across primate genomes.

Gene	Position	SNP	Old World Primate						New World Primate			Lemurs					
			Human			Denisova	Neandertal	Chimp	Gorilla	Orangutan	Baboon	Rhesus	Marmoset	Tarsier	Mouse lemur	Bushbaby	
Alleles			AFA	M.A.F	ASN	M.A.F	EUR	M.A.F									
<i>NCAM1</i>	Intron 13	rs646558	A/C	0,575 (A)	0,196 (A)	A	0,252 (A)	A	A	A	A	A	A	G	G	G	G
<i>TTC12</i>	Exon 3	rs723077*	A/C	0,097 (C)	0,308 (C)	A	0,487 (C)	A	?	A	G	G	G	G	G	A	A
<i>TTC12</i>	Intron 7	rs2303380	A/G	0,338 (G)	0,39 (G)	A	0,381 (G)	A	?	A	A	A	A	A	A	A	A
<i>ANKK1</i>	Exon 8	rs2734849*	A/G	0,115 (G)	0,024 (G)	A	0,473 (G)	A	A	A	A	A	A	A	G	A	=
<i>ANKK1</i>	Exon 8	rs1800497* (Taq1A)	A/G	0,411 (A)	0,407 (A)	A/G	0,195 (A)	A	G	A	A	A	A	A	A	A	=
<i>DRD2</i>	Exon 7	rs6277	G/A	0,034 (A)	0,049 (A)	G	0,534 (A)	G	G/A	G	A	A	A	A	A	G	G
<i>DRD2</i>	Intron 5	rs2283265	C/A	0,092 (A)	0,422 (A)	C	0,167 (A)	C	?	C	C	C	C	C	?	C	=

\*Non-synonymous polymorphism. M.A.F = Minor Allele Frequency, indicated in parenthesis, obtained in HAPMAP. Mean frequencies were computed using European (CEU), Asian (CHB and JPT) and African (YRI) populations. Genomes included: Human (*Homo sapiens*), Denisova (Denisova cave specimen), Neandertal (*Homo neanderthalensis*), Chimpanzee (*Pan troglodytes*), Gorilla (*Gorilla gorilla gorilla*), Orangutan (*Pongo pygmaeus abelii*), Baboon (*Papio hamadryas*), Rhesus (*Macaca mulatta*), Marmoset (*Callithrix jacchus*), Tarsier (*Tarsier syrichta*), Mouse lemur (*Microcebus murinus*) and Bushbaby (*Otolemur garnettii*). (=): aligning species have one or more unalignable bases in the gap region. (?): No information available.

denotes that the G → A mutation possibly occurred on the phylogenetic branch leading to the Old World primates.

The additional analysis of the three non-synonymous SNPs rendered only the Met73Leu (rs723077) mutation in the *TTC12* gene as a non-neutral amino acid change, which the SNAP prediction tool indicated that the protein's chemico-physical properties may change in response to this substitution. The failure in detecting similar results for the other two mutations does not necessarily imply that they have no functional impact since the SNAP, PolyPhen2 and Grantham Score programs were designed to indicate only significant chemical changes in protein structure. Likewise, detecting a non-neutral signal does not necessarily imply direct association with relevant phenotypic changes.

### Orthology, synteny and neighborhood status of the NTAD cluster in vertebrate genomes

Our comparative analysis involved human NTAD gene sequences as query for BLAST searches in the other 46 available vertebrate genome sequences (Table S1). Most of the orthologous genes were previously annotated in the databases used. Additionally, we were able to identify some new *ANKK1* orthologues. This included an unannotated ~1.5kb long *ANKK1* ortholog sequence in the porcine (*Sus scrofa*) genome (Table S1), comprising three exons (orthologous to the human exons 3, 4 and 5; 76% identity), from which a 119 amino acid-long protein sequence could be predicted (86% identity). This *ANKK1* orthologue in pigs is located downstream from *DRD2*, changing the gene order in the cluster from NTAD to NTDA (*NCAM1-TTC12-DRD2-ANKK1*). This interesting result illustrates a unique chromosomal inversion in the porcine lineage (Figure 1). Two other probable new *ANKK1* orthologues were identified in the alpaca (*Vicugna pacos*) and rabbit (*Oryctolagus cuniculus*) genomes (Table S1), but in both cases the expected NTAD gene order was retained.

Conserved synteny and neighborhood of the whole NTAD cluster was observed in 22 of the 48 vertebrate species studied (46%), other 4 species present conserved synteny but do not share full neighborhood (Teleostei: *Danio rerio*, *Gasterosteus aculeatus*, *Tetraodon nigroviridis*; Artiodactyla: *Sus scrofa*; Figure 1). The synteny and neighborhood status of the NTAD cluster in the other genomes could not be ascertained due to low coverage and/or incomplete assembly.

One of the main features of the NTAD cluster (Figure S1) is that *DRD2* is located in the reverse strand (or minus strand) in relation to the other three genes in all vertebrate genomes, except in the porcine genome due to the inversion described above. The presence of more than one *NCAM1* and *DRD2* orthologue copies in teleost fishes (Figure 1B, Table S1) is in agreement with the known genome duplication in this taxon (Jaillon *et al.*, 2009). However, we did not observe conserved synteny in all copies of *NCAM1* and *DRD2*, since the extant copies not included in

the NTAD cluster are located in different chromosomes (Table S1). This raises the question of possible subfunctionalization or neofunctionalization of these gene copies.

It is worthy of note that NTAD cluster synteny is conserved in teleost fishes, but not the neighborhood of its genes. About three million bases, containing ~100 genes, separate *NCAM1* from the TAD cluster (*TTC12-ANKK1-DRD2*), which can be tracked back to at least ~525 Mya in the origin of vertebrates. The *NCAM1* neighborhood was apparently gained when the Sarcopterygians emerged ~400 Mya and seems to have been maintained since then (Figure 1C).

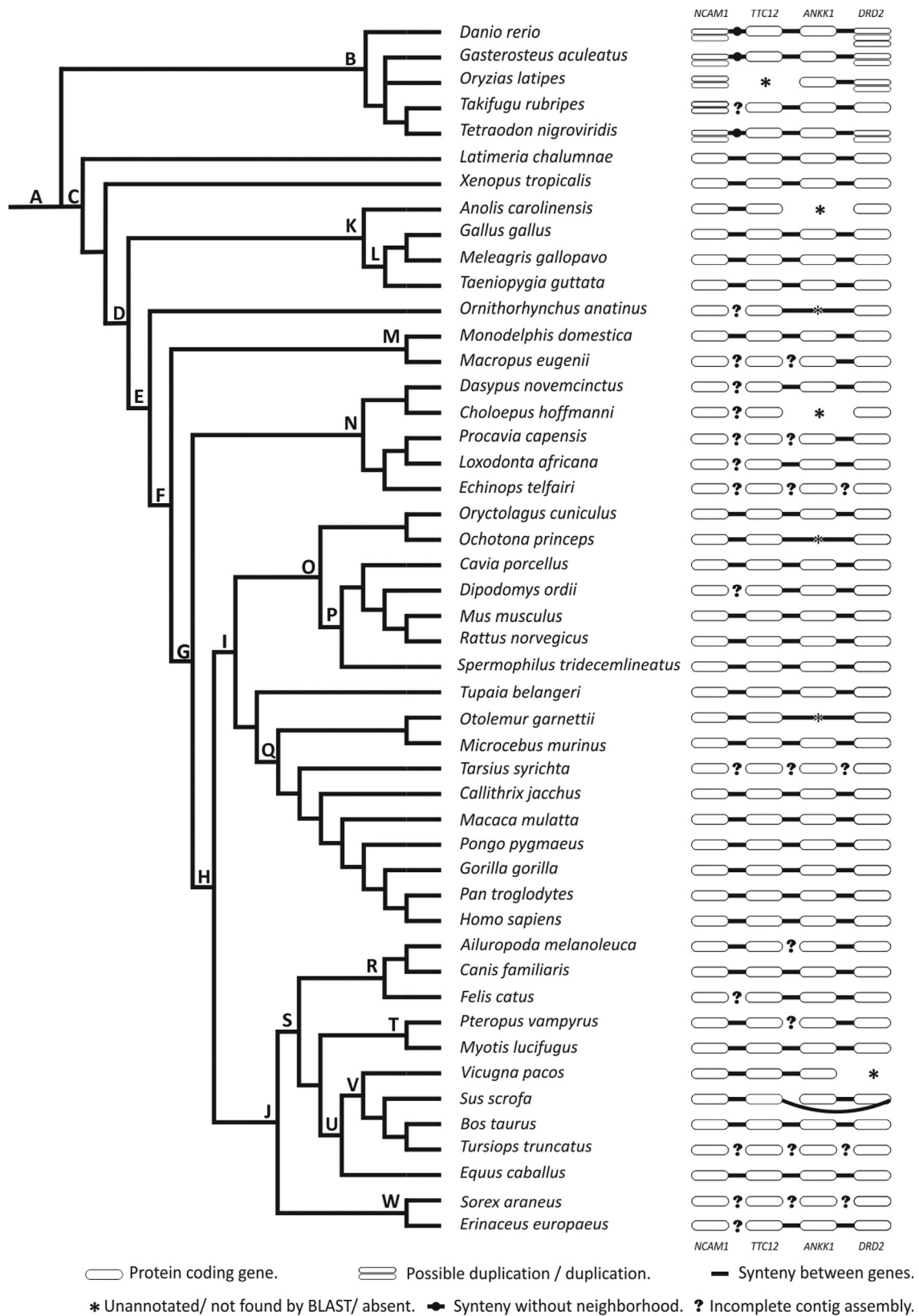
Although phenomena such as tandem duplications, inversions, rearrangements and indels may account for non-random patterns of the genome (Hurst *et al.*, 2004), none of these explain by themselves the origin and conservation of the NTAD cluster. Thus other hypotheses need to be considered.

Several authors have demonstrated that clustered genes are kept together for a long period of time to preserve intact their co-regulatory system and consequently phenotype integrity. For instance, several *cis*-regulatory sequences are preserved throughout all vertebrate genomes due to their role in development (Kikuta *et al.*, 2007). One interesting fact about the NTAD gene cluster is that a polymorphism in one gene might indirectly affect the expression levels of a neighboring gene. Huang *et al.* (2008) demonstrated that the *ANKK1* rs2734849 SNP alters the expression level of NF-κB-regulated genes and, since *DRD2* gene expression is regulated by the transcription factor NF-κB (Fiorentini *et al.*, 2002; Bontempi *et al.*, 2007), it might be indirectly regulated by *ANKK1*. The search for regulatory transcription factors in the ENCODE database showed that the *NCAM1* gene also seems to be regulated by NF-κB, as well as *TTC12* paralogues (Table S2), denoting a possible role of this transcription factor in the co-regulation of the NTAD genes.

Based on computer simulations, Yerushalmi and Teicher (2007) showed an extraordinary tendency for essential genes to cluster as a result of natural selection pressures. Our results illustrate, for the first time to our knowledge, such a tendency for genes with essential functions in neurogenesis and dopaminergic neurotransmission. Thus, it is likely that natural selection, through the formation of the NTAD cluster, has played a role in the emergence of an efficient mechanism of co-regulation, when the vertebrate central nervous system acquired novel traits and gained complexity. Equally important must have been the role of natural selection in maintaining the NTAD cluster practically intact for at least 400 million years.

Notwithstanding, certain limitations must be considered when interpreting the results of the present study. Several assumptions presented here rely on available genomic data and analyses that are still preliminary. This is especially true for tests considering alternative explanations for





**Figure 1** - Conservation of synteny and neighborhood in the NTAD cluster across 48 vertebrate genomes. The tree was compiled based on Agnarsson *et al.* 2010, Asher and Helgen 2010, Horner *et al.* 2007, Janes *et al.* 2010, Page and Goodman 2001, Pacheco *et al.* 2011. A: Vertebrata; B: Teleostei; C: Sarcopterygii; D: Amniota; E: Mammalia; F: Theria; G: Eutheria; H: Boreoeutheria; I: Euarchontoglires; J: Laurasiatheria; K: Sauria; L: Aves; M: Marsupialia; N: Atlantogenata; O: Glires; P: Rodentia; Q: Primates; R: Carnivora; S: Scrotifera; T: Chiroptera; U: Euungulata; V: Artiodactyla; W: Lipotyphla.

specific SNPs that are widespread in the *Homo genus* and for the rate of maintained of four-gene clusters since or before the origin of vertebrates. We understand that a major goal of this study is to stimulate further research in the evolutionary history of gene clusters using the currently available genomic data and emerging bioinformatics tools.

## Conclusion

Our results suggest that genes related to neurogenesis and dopaminergic neurotransmission may be interconnected in the course of the evolution of the complex vertebrate neural system via a common functional genomic architecture and chromosomal dynamics. Associated with due consideration of linkage disequilibrium patterns, this denotes the importance of approaching the NTAD cluster as a candidate functional unit, rather than its genes separately, in behavioral and psychiatric genetic studies.

## Acknowledgments

We thank Brazilian funding agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Conselho de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo a Pesquisa do Rio Grande do Sul (FAPERGS) and Programa de Apoio a Núcleos de Excelência (PRONEX) for financial support.

## References

- Agnarsson I, Kuntner M and May-Collado LJ (2010) Dogs, cats, and kin: A molecular species-level phylogeny of Carnivora. *Mol Phylogenet Evol* 54:726-745.
- Asher RJ and Helgen KM (2010) Nomenclature and placental mammal phylogeny. *BMC Evol Biol* 10:e102.
- Bontempi S, Fiorentini C, Busi C, Guerra N, Spano P and Missale C (2007) Identification and characterization of two nuclear factor-kappaB sites in the regulatory region of the dopamine D2 receptor. *Endocrinology* 148:2563-2570.
- Castelo-Branco G and Arenas E (2006) Function of Wnts in dopaminergic neuron development. *Neurodegener Dis* 3:5-11.
- Ding YC, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK, Flodman P, Spence MA, Schuck S, Swanson JM, *et al.* (2002) Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proc Natl Acad Sci USA* 99:309-314.
- Dubertret C, Bardel C, Ramoz N, Martin PM, Deybach JC, Adès J, Gorwood P and Gouya L (2010) A genetic schizophrenia-susceptibility region located between the *ANKK1* and *DRD2* genes. *Prog Neuropsychopharmacol Biol Psychiatry* 34:492-499.
- Fiorentini C, Guerra N, Facchetti M, Finardi A, Tiberio L, Schiaffonati L, Spano P and Missale C (2002) Nerve growth factor regulates dopamine D(2) receptor expression in prolactinoma cell lines via p75(NGFR)-mediated activation of nuclear factor-kappaB. *Mol Endocrinol* 16:353-366.
- Gelernter J, Yu Y, Weiss R, Brady K, Panhuysen C, Yang BZ, Kranzler HR and Farrer L (2006) Haplotype spanning *TTC12* and *ANKK1*, flanked by the *DRD2* and *NCAMI* loci, is strongly associated to nicotine dependence in two distinct American populations. *Hum Mol Genet* 15:3498-3507.
- Grantham R (1974) Amino acid difference formula to help explain protein evolution. *Science* 185:862-864.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MHY, *et al.* (2010) A draft sequence of the Neandertal genome. *Science* 328:710-722.
- Hall TA (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95-98.
- Horner DS, Lefkimmatis K, Reyes A, Gissi C, Saccone C and Pesole G (2007) Phylogenetic analyses of complete mitochondrial genome sequences suggest a basal divergence of the enigmatic rodent *Anomalurus*. *BMC Evol Biol* 7:e16.
- Huang W, Payne TJ, Ma JZ, Beuten J, Dupont RT, Inohara N and Li MD (2008) Significant association of *ANKK1* and detection of a functional polymorphism with nicotine dependence in an African-American sample. *Neuropsychopharmacology* 34:319-330.
- Hurst LD, Pál C and Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5:299-310.
- Jaillon O, Aury JM and Wincker P (2009) "Changing by doubling", the impact of Whole Genome Duplications in the evolution of eukaryotes. *C R Biol* 332:241-253.
- Janes DE, Organ CL, Fujita MK, Shedlock AM and Edwards SV (2010) Genome evolution in Reptilia, the sister group of mammals. *Annu Rev Genomics Hum Genet* 11:239-264.
- Katoh M and Katoh M (2003) Identification and characterization of *TPARM* gene *in silico*. *Int J Oncol* 23:1213-1217.
- Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engström PG, Fredman D, Akalin A, Caccamo M, Sealy I, Howe K, *et al.* (2007) Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res* 17:545-555.
- Le Foll B, Gallo A, Le Strat Y, Lu L and Gorwood P (2009) Genetics of dopamine receptors and drug addiction: A comprehensive review. *Behav Pharmacol* 20:1-17.
- Li MD, Ma JZ and Beuten J (2004) Progress in searching for susceptibility loci and genes for smoking-related behavior. *Clin Genet* 66:382-392.
- Li WH, Wu CI and Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol* 2:150-174.
- McIntyre JC, Titlow WB and McClintock TS (2010) Axon growth and guidance genes identify nascent, immature, and mature olfactory sensory neurons. *J Neurosci Res* 88:3243-3256.
- Michalak P (2008) Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* 91:243-248.
- Munafò M, Clark T, Johnstone E, Murphy M and Walton R (2004) The genetic basis for smoking behavior: A systematic review and meta-analysis. *Nicotine Tob Res* 6:583-597.
- Munafò MR, Matheson IJ and Flint J (2007) Association of the *DRD2* gene Taq1A polymorphism and alcoholism: A meta-analysis of case-control studies and evidence of publication bias. *Mol Psychiatry* 12:454-461.
- Neville MJ, Johnstone EC and Walton RT (2004) Identification and characterization of *ANKK1*: A novel kinase gene closely

- linked to *DRD2* on chromosome band 11q23.1. *Hum Mutat* 23:540-545.
- Pacheco MA, Battistuzzi FU, Lentino M, Aguilar R, Kumar S and Escalante AA (2011) Evolution of modern birds revealed by mitogenomics: Timing the radiation and origin of major orders. *Mol Biol Evol* 28:1927-1942.
- Page SL and Goodman M (2001) Catarrhine phylogeny: Non-coding DNA evidence for a diphyletic origin of the mangabeys and for a human-chimpanzee clade. *Mol Phylogenet Evol* 18:14-25.
- Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, *et al.* (2011) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468:1053-1060.
- Sémon M and Duret L (2006) Evolutionary origin and maintenance of coexpressed gene clusters in mammals. *Mol Biol Evol* 23:1715-1723.
- Slatkin M and Rannala B (2000) Estimating allele age. *Annu Rev Genomics Hum Genet* 1:225-249.
- Smith L, Watson M, Gates S, Ball D and Foxcroft D (2008) Meta-analysis of the association of the Taq1A polymorphism with the risk of alcohol dependency: A HuGE gene-disease association review. *Am J Epidemiol* 167:125-138.
- Tovo-Rodrigues L, Rohde LA, Roman T, Schmitz M, Polanczyk G, Zeni C, Marques FZ, Contini V, Grevet EH, Belmonte-de-Abreu P, *et al.* (2011) Is there a role for rare variants in *DRD4* gene in the susceptibility for ADHD? Searching for an effect of allelic heterogeneity. *Mol Psychiatry* 17:520-526.
- Wang E, Ding YC, Flodman P, Kidd JR, Kidd KK, Grady DL, Ryder OA, Spence MA, Swanson JM and Moyzis RK (2004) The genetic architecture of selection at the human dopamine receptor D4 (*DRD4*) gene locus. *Am J Hum Genet* 74:931-944.
- Yang BZ, Kranzler HR, Zhao H, Gruen JR, Luo X and Gelernter J (2007) Association of haplotypic variants in *DRD2*, *ANKKI*, *TTC12* and *NCAM1* to alcohol dependence in independent case-control and family samples. *Hum Mol Genet* 16:2844-2853.
- Yerushalmi U and Teicher M (2007) Examining emergence of functional gene clustering in a simulated evolution. *Bull Math Biol* 69:2261-2280.

## Internet Resources

- CisRED: Database of genome-wide regulatory module and element predictions, <http://www.cisred.org/> (June 20, 2011).
- DECODE: Decipherment of DNA Elements, <http://www.sabiosciences.com/chipqpcrsearch.php?app=T> FBS (July 4, 2011).
- Ensembl Genome Browser, <http://www.ensembl.org/index.html> (July 5, 2011).
- GeneCards v.3: the Human Genome Compendium, <http://www.genecards.org/> (May 17, 2011).
- UCSC Genome Browser, <http://genome.ucsc.edu/> (July 10, 2011).
- UniProt Protein Database, <http://www.uniprot.org/> (April 5, 2011).
- PolyPhen2: Prediction of Functional Effects of Human nsSNPs, <http://genetics.bwh.harvard.edu/pph2/> (July 12, 2011).
- SNAP: Effects of Single Amino Acid Substitutions on Protein Function, <http://roslab.org/services/snap/>, (July 12, 2011).

## Supplementary Material

The following online material is available for this article:

- Table S1 - Orthologues of genes in the NTAD cluster in 48 available vertebrate genomes.
- Table S2 - Transcription factors predicted by DECODE as regulators of the NTAD genes in humans.
- Figure S1 - NTAD cluster in the human genome.

This material is available as part of the online article from <http://www.scielo.br/gmb>.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table S1. Orthologues of genes in the NTAD cluster in 47 available vertebrate genomes, showing accession number and location.

Taxon	Genome	Coverage	NCAM1	TTC12	ANKK1	DRD2
TELEOSTEI	<i>Danio rerio</i> (Zebrafish)	~7.5X	ncam1a <a href="#">ENSDARG00000056181</a> <a href="#">21:22615780-23024668:-1</a>	<a href="#">ENSDARG00000056896</a> <a href="#">15:20908817-20942192:1</a>	<a href="#">ENSDARG00000056921</a> <a href="#">15:20975671-20986596:1</a>	drd2a <a href="#">ENSDARG00000056926</a> <a href="#">15:21008646-21030749:-1</a>
			ncam1b <a href="#">ENSDARG00000007220</a> <a href="#">15:17312414-17518321:-1</a>			drd2b <a href="#">ENSDARG00000011091</a> <a href="#">5:59424412-59461143:-1</a>
	<i>Gasterosteus aculeatus</i> (Stickleback)	6X	1 of 2 <a href="#">ENSGACG00000020773</a> <a href="#">groupVII:24387904-24556179:-1</a>	<a href="#">ENSGACG00000009165</a> <a href="#">groupI:8558661-8571918:-1</a>	<a href="#">ENSGACG00000009158</a> <a href="#">groupI:8533757-8538878:-1</a>	drd2 <a href="#">ENSGACG00000009131</a> <a href="#">groupI:8509579-8522810:1</a>
			2 of 2 <a href="#">ENSGACG00000010697</a> <a href="#">groupI:11344900-11377049:1</a>			prediction <a href="#">ENSGACG00000012513</a> <a href="#">groupXX:14002115-14008728:1</a>
<i>Oryzias latipes</i> (Medaka)	10.6X	1 of 2 <a href="#">ENSORLG00000014681</a> <a href="#">14:29382768-29464062:-1</a>	absent	<a href="#">ENSORLG00000007503</a> <a href="#">13:15062437-15069570:1</a>	drd2 <a href="#">ENSORLG00000007515</a> <a href="#">13:15093829-15118346:-1</a>	
		2 of 2 <a href="#">ENSORLG00000005127</a> <a href="#">13:10809441-10820108:-1</a>			prediction <a href="#">ENSORLG00000015196</a> <a href="#">16:22530038-22544806:1</a>	
<i>Takifugu rubripes</i> (Fugu)		8.5X	1 of 2 <a href="#">ENSTRUG00000002966</a> <a href="#">scaffold 454:35293-71171:-1</a>	<a href="#">ENSTRUG00000014546</a> <a href="#">scaffold 9:1673315-1683905:1</a>	<a href="#">ENSTRUG00000014611</a> <a href="#">scaffold 9:1700087-1707491:1</a>	<a href="#">ENSTRUG00000014690</a> <a href="#">scaffold 9:1719056-1732814:-1</a>
			2 of 2 <a href="#">ENSTRUG00000007164</a> <a href="#">scaffold 156:454454-475510:-1</a>			



Table S1 (cont). Orthologues of genes in the NTAD cluster in 47 available vertebrate genomes, showing accession number and location.

Taxon	Genome	Coverage	NCAM1	TTC12	ANKK1	DRD2
TELEOSTEI	<i>Tetraodon nigroviridis</i> (Tetraodon)	7.9X	1 of 2 <a href="#">ENSTNIG00000003848</a> <a href="#">Un_random:27638663-27656353:-1</a>	<a href="#">ENSTNIG00000001367</a> <a href="#">16:938567-948048:-1</a>	<a href="#">ENSTNIG00000007778</a> <a href="#">16:921218-925980:-1</a>	drd2 <a href="#">ENSTNIG00000007779</a> <a href="#">16:901459-913203:1</a>
			2 of 2 <a href="#">ENSTNIG00000016188</a> <a href="#">16:3055810-3078332:1</a>			prediction <a href="#">ENSTNIG00000004342</a> <a href="#">Un_random:68441708-68447001:-1</a>
COELACANTHIMORPHA	<i>Latimeria chalumnae</i>	~7X	<a href="#">ENSLACG00000018069</a> <a href="#">JH126571.1:2687432-2874385</a>	<a href="#">ENSLACG00000018234</a> <a href="#">JH126571.1:2926015-2964315</a>	<a href="#">ENSLACG00000018279</a> <a href="#">JH126571.1:2974006-2990122</a>	<a href="#">ENSLACG00000018381</a> <a href="#">JH126571.1:3151137-3199198</a>
ANURA	<i>Xenopus tropicalis</i>	7.65X	<a href="#">ENSXETG00000011665</a> <a href="#">GL172954.1:1273359-1333515:-1</a>	<a href="#">ENSXETG00000011664</a> <a href="#">GL172954.1:1165658-1229867:-1</a>	<a href="#">ENSXETG00000011663</a> <a href="#">GL172954.1:1081038-1104804:-1</a>	<a href="#">ENSXETG00000011662</a> <a href="#">GL172954.1:921005-972728:1</a>
SAURIA	<i>Anolis carolinensis</i> (Anole Lizard)	7.1X	<a href="#">ENSACAG00000008287</a> <a href="#">GL343354.1:608176-677439:-1</a>	<a href="#">ENSACAG00000008658</a> <a href="#">GL343354.1:553247-588419:-1</a>	absent -	<a href="#">ENSACAG00000008768</a> <a href="#">GL343354.1:451491-460712:1</a> -
			-	-	-	-
	<i>Gallus gallus</i> (chicken)	6.6X	<a href="#">ENSGALG00000007839</a> <a href="#">24:5933204-6012150:-1</a>	<a href="#">ENSGALG00000007833</a> <a href="#">24:5913010-5927569:-1</a>	ANKK1 <a href="#">ENSGALG00000007815</a> <a href="#">24:5895790-5901363:-1</a> LOC415708 (possible ortholog) <a href="#">ENSGALG00000021325</a> <a href="#">11:2632306-2645774:1</a>	<a href="#">ENSGALG00000007794</a> <a href="#">24:5874347-5879422:1</a>
			<a href="#">ENSMGAG00000004316</a> <a href="#">26:6151710-6193006:-1</a>	<a href="#">ENSMGAG00000004280</a> <a href="#">26:6115795-6128289:-1</a>	<a href="#">ENSMGAG00000004269</a> <a href="#">26:6094781-6100367:-1</a>	<a href="#">ENSMGAG00000004223</a> <a href="#">26:6073432-6078559:1</a>
<i>Meleagris gallopavo</i> (Turkey)	17X					
<i>Taeniopygia guttata</i> (Zebra Finch)	6X	<a href="#">ENSTGUG00000000275</a> <a href="#">24:1260906-1302411:-1</a>	<a href="#">ENSTGUG00000000270</a> <a href="#">24:1238535-1250741:-1</a>	ANKK1 <a href="#">ENSTGUG00000000267</a> <a href="#">24:1217973-1221928:-1</a> prediction <a href="#">ENSTGUG00000006862</a> <a href="#">11:6734086-6737862:1</a>	<a href="#">ENSTGUG00000000255</a> <a href="#">24:1191855-1198991:1</a>	

Table S1 (cont). Orthologues of genes in the NTAD cluster in 47 available vertebrate genomes, showing accession number and location.

Taxon	Genome	Coverage	NCAM1	TTC12	ANKK1	DRD2
MONOTREMATA	Ornithorhynchus anatinus (Platypus)	6X	NCAM1 <a href="#">ENSOANG00000009247</a> <a href="#">Contig13035:10432-20417:1</a>	<a href="#">ENSOANG00000010971</a> <a href="#">Contig8462:8719-29891:-1</a>	absent	<a href="#">ENSOANG00000010970</a> <a href="#">Contig8462:1327-6313:1</a>
MARSUPIALIA	Monodelphis domestica (Opossum)	6.8X	<a href="#">ENSMODG00000014314</a> <a href="#">4:239164918-239249671:-1</a>	<a href="#">ENSMODG00000014228</a> <a href="#">4:239070356-239121360:-1</a>	<a href="#">ENSMODG00000014205</a> <a href="#">4:239028762-239043233:-1</a>	<a href="#">ENSMODG00000014196</a> <a href="#">4:238978598-239009228:1</a>
	Macropus eugenii (Wallaby)	2X	<a href="#">ENSMEUG00000012605</a> <a href="#">GeneScaffold_5807:10353-87553:1</a>	<a href="#">ENSMEUG00000009795</a> <a href="#">GeneScaffold_5805:596-50221:1</a>	<a href="#">ENSMEUG00000000701</a> <a href="#">GeneScaffold_5806:4947-18128:1</a>	<a href="#">ENSMEUG00000000707</a> <a href="#">GeneScaffold_5806:34377-57516:-1</a>
XENARTHRA	Dasylops novemcinctus (Armadillo)	2X	<a href="#">ENSNDNOG00000009390</a> <a href="#">GeneScaffold_4191:71042-134109:1</a>	<a href="#">ENSNDNOG00000024773</a> <a href="#">GeneScaffold_4190:3973-86588:1</a>	<a href="#">ENSNDNOG00000013438</a> <a href="#">GeneScaffold_4190:113028-126154:1</a>	<a href="#">ENSNDNOG00000007008</a> <a href="#">GeneScaffold_4190:138117-150433:-1</a>
	Choloepus hoffmanni (Sloth)	2.05X	<a href="#">ENSCHOG00000002918</a> <a href="#">GeneScaffold_4607:4688-64734:1</a>	<a href="#">ENSCHOG00000010881</a> <a href="#">GeneScaffold_4606:474-34240:1</a>	absent	<a href="#">ENSCHOG00000008200</a> <a href="#">GeneScaffold_4608:898-13027:-1</a>
LAGOMORPHA	Oryctolagus cuniculus (Rabbit)	7.48X	<a href="#">ENSOCUG00000006190</a> <a href="#">1:103078226-103142725:-1</a>	<a href="#">ENSOCUG00000011625</a> <a href="#">1:103004607-103049887:-1</a>	blat UCSC (85% id.) chr1:102,979,282-102,990,588	<a href="#">ENSOCUG00000011657</a> <a href="#">1:102961300-102971002:1</a>
	Ochotona princeps (Pika)	1.93X	<a href="#">ENSOPRG00000014977</a> <a href="#">GeneScaffold_2905:194352-257089:1</a>	<a href="#">ENSOPRG00000015703</a> <a href="#">GeneScaffold_2905:281538-325574:1</a>	absent	<a href="#">ENSOPRG00000015750</a> <a href="#">GeneScaffold_2905:347086-356689:-1</a>
RODENTIA	Cavia porcellus (Guinea pig)	6.79X	<a href="#">ENSCPOG00000004681</a> <a href="#">scaffold_19:34540713-34601476:-1</a>	<a href="#">ENSCPOG00000004685</a> <a href="#">scaffold_19:34455258-34511005:-1</a>	<a href="#">ENSCPOG00000001170</a> <a href="#">scaffold_19:34420990-34430699:-1</a>	<a href="#">ENSCPOG00000009265</a> <a href="#">scaffold_19:34398884-34411581:1</a>
	Dipodomys ordii (Kangaroo rat)	1.85X	<a href="#">ENSNDORG00000015038</a> <a href="#">scaffold_2068:5606-68849:-1</a>	<a href="#">ENSNDORG00000013510</a> <a href="#">GeneScaffold_3885:3227-33111:1</a>	<a href="#">ENSNDORG00000013515</a> <a href="#">GeneScaffold_3885:49442-58538:1</a>	<a href="#">ENSNDORG00000013516</a> <a href="#">GeneScaffold_3885:66648-74803:-1</a>
	Mus musculus (Mouse)	high	<a href="#">ENSMUSG00000039542</a> <a href="#">9:49310257-49607027:-1</a>	<a href="#">ENSMUSG00000040219</a> <a href="#">9:49245068-49294330:-1</a>	<a href="#">ENSMUSG00000032257</a> <a href="#">9:49223327-49235126:-1</a>	<a href="#">ENSMUSG00000032259</a> <a href="#">9:49148732-49216282:1</a>
	Rattus norvegicus (Rat)	~2X	<a href="#">ENSRNOG00000031890</a> <a href="#">8:52822361-52885628:-1</a>	<a href="#">ENSRNOG00000008595</a> <a href="#">8:52734796-52799135:-1</a>	<a href="#">ENSRNOG00000025037</a> <a href="#">8:52714745-52722913:-1</a>	<a href="#">ENSRNOG00000008428</a> <a href="#">8:52641169-52707749:1</a>
	Spermophilus tridecemlineatus (Squirrel)	1.90X	<a href="#">ENSSTOG00000003495</a> <a href="#">GeneScaffold_3272:15324-20843:1</a>	<a href="#">ENSSTOG00000003513</a> <a href="#">GeneScaffold_3272:98637-146451:1</a>	<a href="#">ENSSTOG00000003546</a> <a href="#">GeneScaffold_3272:160571-171793:1</a>	<a href="#">ENSSTOG00000003570</a> <a href="#">GeneScaffold_3272:179831-197728:-1</a>

Table S1 (cont). Orthologues of genes in the NTAD cluster in 47 available vertebrate genomes, showing accession number and location.

Taxon	Genome	Coverage	<i>NCAM1</i>	<i>TTC12</i>	<i>ANKK1</i>	<i>DRD2</i>
INSECTIVORA	<i>Sorex araneus</i> (Shrew)	1.9X	<a href="#">ENSSARG00000006760</a> <a href="#">scaffold_255145:60004-73832:-1</a>	<a href="#">ENSSARG00000005049</a> <a href="#">GeneScaffold_4137:32399-65702:1</a>	<a href="#">ENSSARG00000003877</a> <a href="#">GeneScaffold_5578:4431-103273:1</a>	<a href="#">ENSSARG00000001799</a> <a href="#">GeneScaffold_4138:3143-94291:-1</a>
	<i>Echinops telfairi</i> (Lesser hedgehog tenrec)	2X	<a href="#">ENSETEG000000010392</a> <a href="#">scaffold_319785:74303-79506:-1</a>	<a href="#">ENSETEG00000009125</a> <a href="#">GeneScaffold_4950:12296-130505:1</a>	<a href="#">ENSETEG00000002513</a> <a href="#">GeneScaffold_6711:8112-56359:1</a>	<a href="#">ENSETEG000000012787</a> <a href="#">scaffold_222770:3-9483:-1</a>
	<i>Erinaceus europaeus</i> (Hedgehog)	1.86X	<a href="#">ENSEEUG00000000446</a> <a href="#">GeneScaffold_5083:215781-233965:1</a>	<a href="#">ENSEEUG00000004523</a> <a href="#">GeneScaffold_5082:985-76059:1</a>	<a href="#">ENSEEUG00000004637</a> <a href="#">GeneScaffold_5082:95143-106027:1</a>	<a href="#">ENSEEUG00000004688</a> <a href="#">GeneScaffold_5082:114423-125550:-1</a>
SCANDENTIA	<i>Tupaia belangeri</i> (Tree Shrew)	2X	<a href="#">ENSTBEG00000002250</a> <a href="#">GeneScaffold_3191:283798-297944:1</a>	<a href="#">ENSTBEG00000003009</a> <a href="#">GeneScaffold_3191:385371-440831:1</a>	<a href="#">ENSTBEG00000004173</a> <a href="#">GeneScaffold_3191:463,541-476,825</a>	<a href="#">ENSTBEG00000005104</a> <a href="#">GeneScaffold_3191:485260-525277:-1</a>
PRIMATES	<i>Tarsius syrichta</i> (Tarsier)	1.82X	<a href="#">ENSTSYG00000009955</a> <a href="#">GeneScaffold_4816:549-67107:1</a>	<a href="#">ENSTSYG00000012003</a> <a href="#">GeneScaffold_4815:897-48462:1</a>	<a href="#">ENSTSYG00000011390</a> <a href="#">GeneScaffold_6508:3144-6436:1</a>	<a href="#">ENSTSYG00000012280</a> <a href="#">GeneScaffold_4817:26-5539:-1</a>
	<i>Otolemur garnettii</i> (Bushbaby)	1.5X	<a href="#">ENSOGAG00000006401</a> <a href="#">GeneScaffold_3006:89395-100429:1</a>	<a href="#">ENSOGAG00000006402</a> <a href="#">GeneScaffold_3006:190790-254568:1</a>	absent	<a href="#">ENSOGAG00000006407</a> <a href="#">GeneScaffold_3006:308141-320255:-1</a>
	<i>Microcebus murinus</i> (Mouse Lemur)	1.93X	<a href="#">ENSMICG00000003994</a> <a href="#">scaffold_568:82754-94374:1</a>	<a href="#">ENSMICG00000004006</a> <a href="#">scaffold_568:171666-232202:1</a>	<a href="#">ENSMICG00000004027</a> <a href="#">scaffold_568:252261-258554:1</a>	<a href="#">ENSMICG00000004037</a> <a href="#">scaffold_568:267022-280232:-1</a>
	<i>Callithrix jacchus</i> (Marmoset)	6X	<a href="#">ENSCJAG00000012342</a> <a href="#">11:10710214-10782569:1</a>	<a href="#">ENSCJAG00000012306</a> <a href="#">11:10812645-10872609:1</a>	<a href="#">ENSCJAG00000012279</a> <a href="#">11:10889021-10904331:1</a>	<a href="#">ENSCJAG00000012246</a> <a href="#">11:10914111-10929399:-1</a>
	<i>Macaca mulatta</i> (Macaque)	5.1X	<a href="#">ENSMMUG00000004688</a> <a href="#">14:111677926-111745516:1</a>	<a href="#">ENSMMUG00000014328</a> <a href="#">14:111778513-111837307:1</a>	<a href="#">ENSMMUG00000014331</a> <a href="#">14:111852166-111865392:1</a>	<a href="#">ENSMMUG00000014334</a> <a href="#">14:111874406-111941325:-1</a>
	<i>Pongo pygmaeus</i> (Orangutan)	6X	<a href="#">ENSPPYG00000003883</a> <a href="#">11:109978390-110064543:1</a>	<a href="#">ENSPPYG00000003884</a> <a href="#">11:110100792-110153360:1</a>	UCSC chr11:110,175,216-110,187,450	<a href="#">ENSPPYG00000003885</a> <a href="#">11:110197370-110212382:-1</a>
	<i>Gorilla gorilla</i>	2.1X	<a href="#">ENSGGOG00000013097</a> <a href="#">11:111259892-111345449:1</a>	<a href="#">ENSGGOG00000003228</a> <a href="#">11:111381685-111435513:1</a>	<a href="#">ENSGGOG00000003237</a> <a href="#">11:111455141-111468472:1</a>	<a href="#">ENSGGOG00000003244</a> <a href="#">11:111477670-111542012:-1</a>
	<i>Pan troglodytes</i> (Chimpanzee)	6X	<a href="#">ENSPTRG00000004291</a> <a href="#">11:111759836-112076094:1</a>	<a href="#">ENSPTRG00000004292</a> <a href="#">11:112109519-112180130:1</a>	<a href="#">ENSPTRG00000004293</a> <a href="#">11:112184597-112196060:1</a>	<a href="#">ENSPTRG00000004295</a> <a href="#">11:112208040-112222288:-1</a>
	<i>Homo sapiens</i>	high	<a href="#">ENSG00000149294</a> <a href="#">11:112831997-113149158:1</a>	<a href="#">ENSG00000149292</a> <a href="#">11:113185251-113254266:1</a>	<a href="#">ENSG00000170209</a> <a href="#">11:113258513-113271140:1</a>	<a href="#">ENSG00000149295</a> <a href="#">11:113280318-113346001:-1</a>

Table S1 (cont). Orthologues of genes in the NTAD cluster in 47 available vertebrate genomes, showing accession number and location.

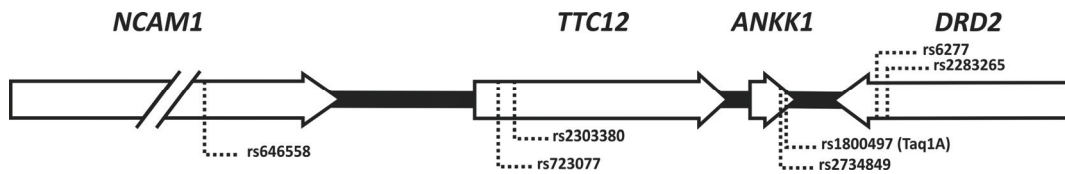
Taxon	Genome	Coverage	<i>NCAM1</i>	<i>TTC12</i>	<i>ANKK1</i>	<i>DRD2</i>
CHIROPTERA	<i>Pteropus vampyrus</i> (Megabat)	2.63X	<a href="#">ENSPVAG00000007824</a> GeneScaffold 2166:404868-466481:1	<a href="#">ENSPVAG00000007826</a> GeneScaffold 2166:495074-539120:1	<a href="#">ENSPVAG00000012852</a> GeneScaffold 2977:967-5576:1	<a href="#">ENSPVAG00000012853</a> GeneScaffold 2977:14934-21721:-1
	<i>Myotis lucifugus</i> (Microbat)	7X	<a href="#">ENSMLUG00000011838</a> GeneScaffold 3345:48657-61150:1	<a href="#">ENSMLUG00000011854</a> GeneScaffold 3345:144154-255563:1	<a href="#">ENSMLUG00000011881</a> GeneScaffold 3345:291777-293760:1	<a href="#">ENSMLUG00000011890</a> GeneScaffold 3345:302226-349693:-1
CARNIVORA	<i>Felis catus</i> (Cat)	1.87X	<a href="#">ENSFCAG00000013891</a> GeneScaffold 2791:232882-301289:1	<a href="#">ENSFCAG00000008291</a> GeneScaffold 2790:17645-63773:1	<a href="#">ENSFCAG00000008292</a> GeneScaffold 2790:84437-101130:1	<a href="#">ENSFCAG00000008293</a> GeneScaffold 2790:109558-121232:-1
	<i>Canis familiaris</i> (Dog)	7.6X	<a href="#">ENSFCAG00000013844</a> 5:22907424-22969273:-1	<a href="#">ENSFCAG00000013902</a> 5:22838477-22881739:-1	<a href="#">ENSFCAG00000013898</a> 5:22810681-22822862:-1	<a href="#">ENSFCAG00000013890</a> 5:22791168-22802738:-1
	<i>Ailuropoda melanoleuca</i> (Panda)	2X	<a href="#">ENSAMEG00000016517</a> GL193553.1:69580-131079:-1	<a href="#">ENSAMEG00000016494</a> GL193553.1:2181-47302:-1	<a href="#">ENSAMEG00000006163</a> GL193846.1:430758-441555:-1	<a href="#">ENSAMEG00000006143</a> GL193846.1:411593-423949:1
ARTIODACTYLA	<i>Bos taurus</i> (Cow)	7.1X	<a href="#">ENSBTAG00000005710</a> 15:21855766-22058994:1	<a href="#">ENSBTAG00000010008</a> 15:22152303-22197195:1	<a href="#">ENSBTAG00000010855</a> 15:22225501-22235297:1	<a href="#">ENSBTAG00000010860</a> 15:22246344-22259339:-1
	<i>Sus scrofa</i> (Pig)	4X	<a href="#">ENSSSCG00000015045</a> 9:39876973-39943731:1	<a href="#">ENSSSCG00000015047</a> 9:39974571-40018618:1	BLAST (80% id) 9:40074571-40076030:-1	<a href="#">ENSSSCG00000015048</a> 9:40051843-40065065:1
	<i>Vicugna pacos</i> (Alpaca)	2.51X	<a href="#">ENSVPAG00000007274</a> GeneScaffold 1967:356795-433389:1	<a href="#">ENSVPAG00000007275</a> GeneScaffold 1967:464956-509310:1	blast (86% id.) GeneScaffold 1967:536135-536217	absent
CETACEA	<i>Tursiops truncatus</i> (Dolphin)	2.59X	<a href="#">ENSTTRG00000000675</a> GeneScaffold 1901:84476-148554:1	<a href="#">ENSTTRG00000010956</a> GeneScaffold 1900:5132-54722:1	<a href="#">ENSTTRG00000014457</a> scaffold 84548:9507-22715:-1	<a href="#">ENSTTRG00000007343</a> scaffold 111984:75732-88099:1
PERISSODACTYLA	<i>Equus caballus</i> (Horse)	6.8X	<a href="#">ENSECAG00000019193</a> 7:21647755-21713496:1	<a href="#">ENSECAG00000017966</a> 7:21745287-21793507:1	<a href="#">ENSECAG00000011781</a> 7:21816607-21822636:1	<a href="#">ENSECAG00000013567</a> 7:21833004-21893006:-1
HYRACOIDEA	<i>Procavia capensis</i> (Hyrax)	2.19X	<a href="#">ENSPCAG00000005740</a> scaffold 1490:10531-76838:-1	<a href="#">ENSPCAG00000016245</a> GeneScaffold 4301:8305-47937:1	<a href="#">ENSPCAG00000006102</a> GeneScaffold 5824:10185-18859:1	<a href="#">ENSPCAG00000006199</a> GeneScaffold 5824:29553-42938:-1
PROBOSCIDEA	<i>Loxodonta africana</i> (Elephant)	7X	<a href="#">ENSLAFG00000005371</a> scaffold 147:1390575-1461424:1	<a href="#">ENSLAFG00000017571</a> scaffold 58:59908-112135:1	<a href="#">ENSLAFG00000025747</a> scaffold 58:151268-165444:1	<a href="#">ENSLAFG00000006523</a> scaffold 58:176422-189953:-1



Table S2. Transcription factors predicted by DECODE as regulators of the NTAD genes in humans.

Gene	Transcription factors
<i>NCAM1</i>	<b>NF-κB</b> , GR-α, GR-β, AP-1, c-Fos, c-Jun, p53, Sox9, POU2F1.
<i>TTC12*</i>	HOXA9B, Meis-1, RFX1, MIF-1, AREB6, Nkx3-1 v4, Nkx3-1, Nkx3-1 v1, Nkx3-1 v2, Nkx3-1 v3.
<i>ANKK1</i>	<b>NF-κB</b> , CUTL1, AREB6, oct-B3, oct-B2, Oct-B1, POU2F2, POU2F2 (Oct-2.1), POU2F1, POU2F2B, POU2F2C.
<i>DRD2</i>	<b>NF-κB</b> , Egr-1, CREB, GR-α, GR-β, δCREB, c-Rel, p300, NRSF form 1.

\*Although NF-κB was not predicted as a regulator of *TTC12*, it regulates the *TTC12* paralogues *TOMM34* and *STIP1* as well as other genes encoding the tetratricopeptide repeat domain (not shown).



**Figure S1-** The NTAD cluster in the human genome, showing the single nucleotide polymorphisms (SNPs) included in this study. The sizes of the four genes are 317 kb, 69 kb, 13 kb and 66 kb, from left to right. Only *DRD2* is positioned in the reverse strand (arrow pointing to the left).