COGNITIVE DOMAINS FUNCTION COMPLEMENTATION

BY NTNG GENE PARALOGS

Pavel Prosselkov\textsuperscript{1,2,*}, Denis Polygalov\textsuperscript{3}, Qi Zhang\textsuperscript{1}, Thomas J. McHugh\textsuperscript{3}, and Shigeyoshi Itohara\textsuperscript{1,*}

\textsuperscript{1}Laboratory of Behavioral Genetics, RIKEN Brain Science Institute, Wakoshi, 351-0198 Saitama, Japan

\textsuperscript{2}Graduate School, Department of Veterinary Medicine, Faculty of Agriculture, Tokyo University, 113-8657 Tokyo, Japan

\textsuperscript{3}Laboratory of Circuit and Behavior Physiology, RIKEN Brain Science Institute, 351-0198 Saitama, Japan

*corresponding authors:

prosselkov@brain.riken.jp

sitohara@brain.riken.jp

Keywords: gene duplication, evolution, subfunctionalisation, cognitive complementation
Gene duplication was proposed by S. Ohno (Ohno, 1970) as a key mechanism of a gene function evolution. A pair of gene paralogs, NTNG1 and NTNG2, sharing identical gene and protein structures and encoding similar proteins, forms a functional complement subfunctionalising (SF) within cognitive domains and forming cognitive endophenotypes, as detected by Intellectual Quotient (IQ) tests (Prosselkov et al., 2015). NTNG paralogs are associated with autism spectrum disorder (ASD), bipolar disorder (BD) and schizophrenia (SCZ), with unique non-overlapping segregation among the other 15 cognitive disorders (CD), emphasizing an evolutionary gain-dependent link between advanced cognitive functions and concomitant neurocognitive pathologies.

Complementary expression and human brain transcriptome composition of the paralogs explains the observed phenomena of their functional complementarity. The lowest identity among NTNGs is found in a middle of encoded by them proteins designated as unknown (Ukd) domain. NTNG1 contains anthropoid-specific constrained regions and both genes contain non-coding conserved sequences underwent accelerated evolution in human. NTNG paralogs SF perturbates “structure drives function” concept at protein and gene levels. Their functions diversification result in a so-called “Cognitive Complement (CC)” formation, a product of gene duplication and subsequent cognitive subfunction bifurcation among NTNG gene duplicates.
Introduction

Gene duplication was proposed by S. Ohno (Ohno, 1970) as a key mechanism of a gene function evolution. Complex behaviors arise from a combination of simpler genetic modules that either have evolved separately or co-evolved. Many genes and the proteins they encode have been found to be involved in the cognitive information processing with a single variant or a single gene generally accounting for only a partial phenotypic variation in a complex trait. Cognitive processing as a quintessence of brain functioning can be viewed as a product of intricately interlinked networks generated by deeply embedded into it players with specific or partially overlapping functions. The robustness of the cognitive processing towards its single elements genetic eliminations (to study their function) and its simultaneous fragility expressed in the multiple forms of neurological disorders manifest the existence of cognitive domains interlocked but SF within a unit of cognition formed upon these domains interaction.

Previously, we have described a function of a pair of gene paralogs (NTNG1 and NTNG2) involved in human IQ tests performance which underwent hominin-specific evolutionary changes (Prosselkov et al., 2015). Hereby, we continue looking at these genes paralogs features focusing on underlying mechanisms of their function segregation and complementation within the cognitive domains.
RESULTS

The previously observed phenomena of functional complementation among the NTNG paralogs within cognitive domains (Prosselkov et al., 2015) is also manifested in NTNG-associated human pathologies diagnosed in most cases (if only not in all) by a cognitive decline (Figure 1A-1 and A-2). Both genes are associated with BD and SCZ – devastating disorders sharing similar etiology (Lee et al., 2013) with genetic correlation by multivariate analysis of 0.590 (Maier et al., 2015), linked to human creativity (Power et al., 2015), and characterized by impulsiveness as a common diagnostic feature (Reddy et al., 2014). Recently found associations of both paralogs with ASD (Sanders et al., 2015) supports the reported genetic correlation of 0.194 ASD/SCZ pair (Maier et al., 2015) and shared module eigengenes detected by PC1 among these two disorders (Parikshak et al., 2015). 12 NTNG1-linked CDs, ranging from AD to TS, span a broad spectrum of clinical features frequently involving reduced processing speed (PS) and verbal comprehension (VC, Figure 1A-1). As for NTNG2, working memory (WM) deficit and inability “to bind” events (perceptual organization, PO) are the most prominent diagnostic traits for the SLE and TLE patients (Figure 1A-2), with PN characterised by indolent behavior in 90% of the cases (Cavard et al., 2009). Thus, both NTNG paralogs are associated with a variety of CDs and mostly in a non-overlapping manner, except for ASD, BD and SCZ characterized by shared and wide spectrum of cognitive abnormalities. The clinical etiology of the aforementioned diseases supports the IQ-deduced functional complementation among the NTNG paralogs (Prosselkov et al., 2015) with (VC/PS) and (WM/PO) deficits being also uniquely segregated among the associated cognitive pathologies.

Since both genes are expected to have identical gene exon/intron compositions but different in their intron lengths (Yin et al., 2002) we have reconstructed the paralogs transcriptomes by re-processing the publicly available RNA-seq dataset (Wu et al., 2012).
from healthy and SCZ human subjects superior temporal gyrus (STG) post-mortem brain tissue (Supplementary Table 1a=ST1a). A difference is noted instantly at the total expression levels (genes, exons, individual RNA transcripts) when two gene paralogs are compared (Figure 1B-1 and B-2). NTNG2 amount (as a whole gene) is 5 times larger comparing to NTNG1; exons (2-5) are 3 times, exons (8-9) are 18 times and exon 10 is 4 times higher expressed for NTNG2 than for NTNG1. The only two exons outlaying the prevailing amount rule for the NTNG2 mRNAs are exons 6 and 7, expressed nearly at the same absolute level as for the NTNG1 exon paralogs, making them highly underrepresented within the whole NTNG2 transcriptome. Next, distinct non-alternating splicing modules are formed by exons (2-5) for NTNG1 (Figure 1B-1), while exons (4-5) and exons (8-9) for NTNG2 (Figure 1B-2). Two structurally identical RNA transcript paralogs (NTNG1a = G1a and NTNG2a = G2a) have been found to exist in both NTNG transcriptomes with G2a being expressed at 8-9 times higher level than G1a. NTNG1 is uniformly presented across the all analysed 16 human samples by 2 more protein coding RNAs (G1c and G1d, detected previously in mice brain, Nakashiba et al., 2000) and by 2 non-coding intron (9-10) derived transcripts (Figure 1B-1). At the same time, NTNG2 transcriptome is comprised of one extra potentially coding RNA (G2a-like with exon 2 spliced out but in-frame coding preserved) and 2 assumed to be non-coding RNAs with exons 6 and 7 retained along with preceding and following them introns. Quite interesting that these two latter transcripts are the only RNA species with NTNG2 exon 6 and 7 retained (Figure 1B-2). Two more coding (G1f and G1n) and 4 more non-coding for NTNG1 and 9 extra non-coding for NTNG2 RNA species have been also assembled from the available reads but due to inconsistency in their appearance across all 16 STG samples they are not presented on the figure but summarized in the table (Figure 1C, for details refer to ST1d). Summarising this, quantitative and qualitative complementary differences is a prominent feature characterising the brain RNA transcriptome of human NTNG paralogs.
However, no significant changes at the transcription level of neither whole genes, nor individual exons, nor reconstructed RNA transcripts have been found for SCZ and healthy subjects.

Upon calling the presence of IQ-affecting SNPs (Prosselkov et al., 2015) across all STG samples (ST1c) it has been revealed that 15 out of 16 subjects were positive for the T-allele of rs2149171 (exon 4-nested), shown above to attenuate the WM score in SCZ patients, making a comparison among the allele carrier vs non-carrier impossible. Four healthy and three SCZ samples carry a T-allele of rs3824574 (exon 3-nested, non-affecting IQ), and 1 healthy and 1 SCZ sample each contains a C-allele of rs4915045 (exon10, non-coding part-nested, and non-affecting IQ). Thus, among the eleven cognitive endophenotype-associated SNPs only 3 were possible to call out of the available NTNG transcriptome.

Distinctly complementary nature of the NTNG paralogs segregation within neurological disorders and RNA transcriptome usage in STG (Figure 1) has prompted us to analyse both genes expression across the entire human brain. We have reconstructed both genes expression profiles in the human brain areas over the life span from conception (pcw = post-conception week) to mature age (30-40 yrs old) using the RNA-seq data from BrainSpan (www.brainspan.org). Similarities and differences are easily noted when the age-dependent phases of NTNG1 and NTNG2 expression profiles are matched (Figure 2). Based on the visual inputs three distinct classifiers have been elaborated: 1. predominantly synchronous (Figure 2A(1-4)), characteristic mostly for the cortical areas; 2. predominantly mixed and asynchronous (Figure 2B), characteristic for the cerebellar cortex and subcortical formations; and 3. anti-phasic (complementary, Figure 2C), characteristic for the MD of thalamus and hippocampus. All analysed brain areas demonstrated an elevated level of NTNG2 expression in comparison to NTNG1 except for thalamus (Figure 2C) with the largest difference observed is at the time of birth (35-37 pcw) or soon after (4 mo) for the synchronous
classifiers (Figure 2A), oscillating increment values across the life span for the mixed (Figure 2B) and anti-phasic (Figure 2C) classifiers. It is quite intriguing to note that essentially all brain areas show a trend towards the expression difference being negated between the paralogs by reaching the mature age of 30-40 yrs old (nearly or above the mean age used for the IQ testing), except MD where the expression discrepancy is increased. Thus, the observed functional complementation among the NTNG paralogs is supported by the anatomical distribution of the genes in human brain and their expression pattern modality over the human subjects lifetime.

A direct comparison of the NTNG paralogs shows not only identical intron-exon gene structure (Figure 1B-1, 2B-2) but also closely matched exon sizes (Figure 3A). There are three exons of identical sizes (exons 4, 8 and 9), another three exons differed by one encoded aa (exons 3, 5 and 6) and there are exons of different sizes (exons 2, 7 and 10). In terms of size the largest difference among the genes is visually presented by the introns: intron (9-10) of NTNG1 is 52.7 times larger its NTNG2 paralogous intron with intron (6-7) of NTNG1 being only 1.43-times larger pointing towards non-equilibria process of non-coding elements elaborations as the process of gene paralogs SF proceeded. Nevertheless, it can be generalised that in average all NTNG1 introns are several times larger their NTNG2 analogs (Figure 3A). We have shown previously that exons 6 and 7 are differentially used within the brain NTNG transcriptome (Figure 1B-1 and B-2) and to explore their potential contribution into the paralogs SF we have built identity matrices with these exons excluded and included (but still producing in-frame existing transcripts, Figure 3B-1 left and right panels, respectively). Exclusion of both exons from the full-lengths transcripts (thus converting NTNG1m to NTNG1a and NTNG2b to NTNG2a, respectively) increases the identity of DNA on 2% (a relatively large effect since both exons together represent only 7.22 and 9.69% of the total coding part of the full-length RNA transcripts, NTNG1m and NTNG2b, respectively).
This effect becomes even stronger when the encoded by these transcripts proteins are also compared (Figure 3B-2). The spliced out Ukd protein domains (encoded by the exons 6 and 7) increases the proteins identity on 3.8% thus making the middle of both genes (and encoded proteins) substantially more different among the both gene paralogs. To corroborate this observation and to explore the importance of other protein parts we have directly compared the sequences encoded by the full-length transcripts and producing Netrin-G1m and Netrin-G2b (Figure 3C). Similarly to what has been shown on Figure 3B-1 and 3B-2, the lowest identity (17.5%) is represented by the Ukd domain (encoded by the exons 6 and 7) and by the preceding exon 5 (a 3’-part of the LE1 domain). Two other areas also show a substantially low identity, namely the N-terminus (it includes the protein secretory signal indicated by an arrow) and the outmost C-terminus responsible for the unique feature of Netrin-Gs – the GPI attachment. Thus, based on the percent identity comparisons among the Netrin-G paralogs it can be predicted that there are several potential protein parts contributing to the paralogs SF. As it has been reported by Seiradake et al. (2011), identical gene and protein domain compositions result in the identical structural motif with differences only in the spatial arrangement of the loops facing the post-synaptic Netrin-G’s interacting partners, NGL-1 and NGL-2, respectively (Figure 3D). Loop I binding surfaces alignment (Figure 3C, blue color) shows a high level of conservation (with at least 5 amino acids 100% conserved) among the Netrin-G paralogs, indicating that it is unlikely to be responsible for the cognate ligand binding specificity. Neither Loop II (Figure 5C, yellow color) nor Loop III (Figure 5C, orange color) display a single conserved amino acid shared among the paralogous binding interfaces (as it originally has been described in Seiradake et al., 2011). Thus the complementary pattern of the pre-postsynaptic interactions mediated via specific Netrin-G/NGL pairs is reflected in the reciprocally different sizes of the loops binding interfaces representing another element of the NTNG-encoded protein paralogs SF.
DISCUSSION

Complementary contribution of NTNG paralogs into human cognitive pathologies.

Involvement of the pre-synaptically expressed axon-localised NTNGs in SCZ diagnosis supports the established view of SCZ as a result of distorted trans-synaptic signaling (Lips et al., 2012), with a recent study proving that axonal connectivity-associated genes form a functional network visualisable by fMRI (Richiardi et al., 2015), and that brain connectivity predicts the level of fluid intelligence (Finn et al., 2015; Pamplona et al., 2015). Both NTNGs have been found to participate in the brain functional connectivity by the parcellated connectome reconstruction (Hawrylycs et al., 2015). Most of the reported disease associations link NTNG1 to SCZ with a variety of other neurologic pathologies (15 in total, Figure 1A-1), while NTNG2 pathologic associations (6 in total, Figure 1A-2) are quite limited to those affecting WM or PO. Among them is SLE frequently characterized by WM deficit (Shucard et al., 2011) and also known to represent schizoid-type abnormalities characteristic for autoimmune pathologies (Guilloux et al., 2010; Eaton et al., 2006).

Immune activation is known to lead to altered pre-pulse inhibition (a key diagnostic trait for SCZ) reversed by antipsychotics (Romero et al., 2007). The three diseases associated with both paralogs (ASD, BD and SCZ) are also a primary focus of the recently initiated PsychENCODE project (PsychENCODE et al., 2015). It is also worth to mention the resemblance of the reported disease associations with the behavioral phenotypes of Ntng1 and Ntng2 gene knockout mice (Zhang et al., 2016).

A gene content associated with IQ score often relates to numerous diseases, such as SCZ, ASD, depression, and others (see Zhao et al., 2014 for ref.; Johnson et al., 2015). Several genes associated with SCZ have undergone positive selection following the human brain evolution (Xu et al., 2015). Despite the global network properties of the brain transcriptome are highly conserved among the species there are robust human-specific
disease-associated modules (Miller et al., 2010) and human accelerated regions (HARs) - highly conserved parts of genome that underwent accelerated evolution in humans (Pollard et al., 2006). HARs can serve as genomic markers for human-specific traits underlying a recent acquisition of modern human cognitive abilities by brain (Boyd et al., 2015) but that also “might have led to an increase in structural instability… resulted in a higher risk for neurodegeneration in the aging brain” (Zhou et al., 2015), rendering our intellectual abilities genetically fragile (Crabtree, 2013) and resulting in a variety of CDs. The role genomic context, epistasis (Hemani et al., 2014), plays in the evolution and pathology is manifested by frequently found disease-causing alleles present in animals without obvious pathological symptoms for the host (Jordan et al., 2015). Any CD is characterized by general intellectual disability (GID) plus psychiatric symptoms. A genetic perturbation-exerted behavioral cognitive deficit (BCD) in an animal model organism is a poor match to a human CD *per se* due to very poor contextual resemblance between the human GID and animal BCD together with the absence of interpretable psychiatric symptoms. Usefulness of animals as psychiatric models is also compromised by the fact that transcriptome differences within species tissues is smaller than among the homologous tissues of different species (Barbosa-Morais et al., 2012; Lin et al., 2014). No wonder that the compounds that “cure” mice models consistently fail in human trials (discussed in Hyman, 2014).

**NTNG paralogs brain transcriptome intrinsic complementarity and possible mechanism for the IQ-affecting mutation alleles effect.** There is no global change at the mRNA level between healthy subjects and SCZ patients (Figure 1B). This conclusion is supported by previously published works stating that globally altered mRNA expression of *NTNG1* or *NTNG2* is unlikely to confer disease susceptibility, at least in the temporal lobe (Eastwood and Harrison, 2008), and Brodmann’s area (Aoki-Suzuki et al., 2005). However, the original paper-source of the STG samples RNA-seq along with many other genes (>1,000)
found that NTNG1 (but not NTNG2) falls under the group of genes with significant alternative promoter usage (Wu et al., 2012: ST6, \(p<9.05\times 10^{-10}\) at FDR <0.5) and NTNG2 (but not NTNG1) clusters with genes (>700) with significant alternative splicing change (Wu et al., 2012: ST7, \(p<6.15\times 10^{-12}\) at FDR <0.5) when SCZ and controls are compared. Such GWAS observation adds an extra layer of complementary regulation to both NTNG paralogs on a top of the described in the results section complementary usage rule for the exons, formed unspliced splicing modules, resulting transcripts and their comprising exons (Figure 1B). Based on the available RNA-seq dataset it was almost impossible to detect RNA with the matched position of NTNG SNPs used for the IQ testing (ST2c) except for two coding exons located (rs2149171 and rs3824574) and exon 10 non-coding area located but transcribed rs4915045 (in 2 out of 16 samples). This fact points towards indirect effect of the IQ-affecting mutation alleles potentially associated with shorter (secretable) isoforms generation (Prosselkov et al., unpublished) lacking two of the most prominent NTNG features: GPI-link and the Ukd domain through an aberrant splicing factor binding. The GPI-link is a hallmark of Netrin-G family members (Nakashiba et al., 2000, 2002) and without it the aberrant Netrin-G isoforms are likely to mimic the action of their releasable ancestry molecules - netrins, still being able to bind to their cognate postsynaptic ligand – NGL but without forming an axonal-postsynaptic contact. The Ukd domain of Netrin-G1, despite its so-far unknown function, is involved in lateral binding to the pre-synaptically localised LAR modulating the binding strength between NGL-1 and Netrin-G1 (Song et al., 2013). Work is currently underway in search for a similar lateral interaction partner for the Netrin-G2 Ukd domain (Kim E, personal communications). The inclusion of Ukd encoding exons 6 and 7 is regulated by the Nova splicing factor (Ule et al., 2005) affecting the cortex Netrin-G1 exon 7 but not exon 6, and, simultaneously, Netrin-G2 paralog exons exhibiting an opposite pattern. In general, it is tempting to speculate that deregulation of NTNG transcripts
processing may have a role in the brain-controlled cognitive abilities and associated CDs.

Supporting such notion, a decreased level of Netrin-G1c mRNA (exons 6-9 excluded, Figure 1B-1) has been reported for BD and SCZ (Eastwood and Harrison, 2008) with Netrin-G1d (exons 6 and 7 included but 8-9 excluded, Figure 1B-1) and Netrin-G1f (a secretable short isoform consisted of domain VI only and lacking the Ukd and GPI-link) being increased in BD, but not in SCZ, in anterior cingulated cortex (Eastwood and Harrison, 2010). Higher Netrin-G1d mRNA expression in fetal brain but low for the Netrin-G1c isoform in the human adult (Eastwood and Harrison, 2008) indicates different functionality of these two splice variants juggling with the Ukd domain inclusion. And, according to our other data, if Netrin-G1 Ukd-containing isoforms are the dominant isoforms in adult mouse brain, Netrin-G2 Ukd-containing isoforms are present only at the trace level (Prosselkov et al., forthcoming), resembling a similar transcriptome pattern for the human STG samples (Figure 1B-1 and B-2). A similar “dynamic microexon regulation” associated with the protein interactome misregulation has been reported to be linked to ASD (Irimia et al., 2014).

Synchronous and complementary expression of NTNG paralogs in the human brain supports the IQ-associated cognitive endophenotypes. Influential parieto-frontal integration theory (P-FIT, Jung and Haier, 2007) states that general intelligence (“g”) is dependent on multiple brain cortical areas such as dlPFC, Broca's and Wernicke's areas, somatosensory and visual cortices (Colom et al., 2009). Despite “g” is widely accepted as the only correlate of the intelligence, its unitary nature was challenged by (Hampshire et al., 2012) claiming had indentified two independent brain networks (for memory and for reasoning) responsible for the task performance, the idea later criticised for the employed data processing approach (Haier et al., 2014). Higher IQ scores (a composite surrogate of “g”) have been reportedly associated with the fronto-parietal network (FPN) connectivity (Song et al., 2008; Glascher et al., 2009). High level of NTNG paralogs expression within
the cognition intensively loaded areas of the brain and the distinct patterns of expression profiles (synchronous, asynchronous/mixed, and complementary, Figure 2A) support associations of NTNG1 and NTNG2 with the recorded cognitive endophenotypes (Prosselkov et al., 2015). Based on the expression patterning over the human life-span, among the total 16 analysed brain areas we found two falling under the same “anti-phasic (complementary)” classifier (Figure 2C): HIP and MD. Adding more to that, MD is the only brain area (out of the 16 presented) where NTNG1 expression level exceeds that of NTNG2 making it a promising candidate for the phenomena of NTNGs SF explanation. Two other brain areas classified by a synchronous paralogs expression deserve a special attention, dIPFC and mPFC (Figure 2A-4). PFC circuitry has been known as a “hub of the brain’s WM system” (Kim et al., 2013; Markowitz et al., 2015), which acts through direct HIP afferents (Spellman et al., 2015) and has many connections with other cortical and subcortical areas (Riga et al., 2014). mPFC may function as an intelligence-control switchboard and IPFC, part of the FPN global connectivity, predicts the WM performance and fluid intelligence (Cole et al., 2012). Interactions of the auditory recognition information fed by the vPFC stream with the sequence processing by the dorsal stream are crucial for the human language articulation (Skeide and Friederici, 2015; Thothathiri and Rattinger, 2015). The fact that both NTNG paralogs are extensively expressed across PFC (Figure 2A-2 and A-4) pinpoints this area as a key for future molecular studies of the human-unique symbolic communications. And PFC is not only implicated in many psychiatric disorders, including SCZ (Gulsuner and McClellan, 2014; see also Riga et al., 2014 for ref.), but is also the only brain structure unique to primates without known homologs in the animal kingdom (Wise, 2008).

Evolution of the protein paralogs encoded by the NTNGs. Forkhead box P2 (FOXP2) – a ubiquitously expressed transcription factor that has been reported to be linked to the evolution of human language through T303N, N325S substitutions when compared to a
primate ortholog (Enard et al., 2002) and is 100% identical to Nea protein (Krause et al., 2007). FOXP2 regulates expression of multiple genes in human and chimpanzee (Konopka et al., 2009), and among them is an M3 gene brain module representative responsible for general fluid cognitive abilities (Johnson et al., 2015), LRRC4C, a gene encoding NGL-1 – a post-synaptic target of Netrin-G1. Similarly to FOXP2, Netrin-G1 is a 100% conserved protein among the hominins with only 1 mutation found in chimpanzee which is absent in marmoset (and other primates) and mice proteins (Prosselkov et al., 2015). On the other hand, extinct hominins’ Netrin-G2 relatively to modern human contains T346A point mutation (as per current version of hg19), also found in primates and mouse and known as rs4962173 (dbSNP missense mutation) representing an ancient substitution from Neandethal genomes found in modern humans and reflecting a recent acquisition of the novel allele around 5,300 yrs BC. Nothing is known regarding the functional significance of this mutation but biochemically a substitution of alanine (A) on a polar threonine (T) could bring an extra point of regulation, e.g. a phosphorylation or glycosylation (NetPhos2.0 (Blom et al., 1999) assigns a low score for the T346 to be phosphorylated but NetOGlyc4.0 (Steentoft et al., 2013) robustly predicts it to be glycosylated, SM). Another mutation S371A/V reflects a selective sweep in Netrin-G2 protein from primates to hominins within a similar to T346A functional context when a hydrophobic alanine (in chimpanzee, A)/valine (in marmoset, V) is replaced by a polar serine (S) and a strong positive predictions for glycosylation but not phosphorylation (SM). This poses a question whether these two human-specific protein substitutions associate with advanced cognitive traits as they may represent a hidden layer of poorly studied so far protein glycosylation-associated regulatome known to affect the brain function and diseases (Baenziger, 2012; Baenziger, 2013). Adding more to this, T346 is nested on exon 5 just 20 nu away from the affecting WM score rs2274855 (Prosselkov et al., 2015), and, together with S371A/V, they are both located within the lowest percent identity.
area (exons (5-7)) of Netrin-Gs (Figure 3C) and, proposedly, contributing to the NTNG duplicates SF. There are at least three more protein parts potentially contributing to the gene paralogs specialised function subdivision (based on the low identity scores, Figure 3C): the secretory peptide, the GPI-link, and the outmost structurally elaborated unstructured loops (I-III) responsible for the reciprocal binding of Netrin-Gs to their post-synaptic cognate partners, NGL-1 or NGL-2, both containing a C-terminal PDZ-binding domain (Kim et al., 2006). An interesting finding was reported in (Arbuckle et al., 2010) found a presence of SH3(PSD95) domain binding site (required for the phosphatidylinositol-3-kinase recruitment) in mice Netrin-G2 (100% identical to human) but not in Netrin-G1. The detected SH3 binding site overlaps with the Netrin-G2-loop III responsible for the binding specificity to NGL-2 (Seiradake et al., 2011; Soto et al., 2013; DeNardo et al., 2012). A plausible working hypothesis would be that while internalised (and being GPI-link naïve/immature) the pre-synaptic Netrin-G2 is bound to SH3-PSD95 via loop III but as soon as being secreted extracellularly (and being attached to the membrane) it is bound to post-synaptic NGL-2. Corroborating this, in the absence of Netrin-G2 in the KO mice NGL-2 is unstable on the post-synaptic surface and gets quickly internalised (Zhang et al., 2016). We can only speculate regarding the potential importance of PSD-95(SH3)-Netrin-G2-NGL-2 scaffolding loop interaction/competition but the ability for Netrin-G1 to bind to SH3 has not been reported. Following this logic, Netrin-G1 should have a similar binding partner via loop II.

The overall identical structural scaffold among the Netrin-G paralogs (Figure 3D) is likely to represent an ancienctly preserved one of the primordial protein (encoded by a single gene in the primitive urochordate C.intestinalis) and its contribution to the process of SF among the NTNG paralogs goes against the “structure drives function” concept. It looks like that it is not the “structure” but rather the “evolution” itself that drives a selection for the best structural (or unstructural in our case) fit out of the available frameworks provided by the
gene duplicates to fulfill the emerged functional demand in a new ecological niche. The intricate variability of phenotype is grounded by the conserved nature of genotype and constrained by the “structure-function” limitations of the coding DNA and is only possible due to permissive evolutionary continuing elaborations of non-coding areas able to absorb the most recently acquired elements (having a potential to become regulatory at some point, e.g. like HAR5 (Boyd et al., 2015)) and carried over by neutral drift as proposed by Kimura but for proteins (Kimura, 1983). At the same time, the multiple protein substitutions coinciding with the SF labor segregation phenomena among the Netrin-G paralogs question their neutral nature. Both of them undergo a purifying selection from mice to human through the reduction in size of non-coding DNA (introns) and encoded proteins (the mice Netrin-G2 is 2 aa longer its human ortholog) further contributing to the host-specific SF. Thus while the non-coding sequences are used to explore the evolutionary space in time, the restrictive boundaries of the paralogs SF are determined by the protein (unstructured) elements.

**Molecular evolution of the Cognitive Complement (CC).** Appearance of the neural crest (Abitua et al., 2012), an event that “affected the chordate evolution in the unprecedented manner” (Green et al., 2015), multipotent progenitor cells (Stolfi et al., 2015), and neurogenic placodes (suggesting a chemosensory and neurosecretory activities, Abitua et al., 2015) in first primitive urochordates/tunicates coincides with the presence of Ntng precursor gene (ENSCING00000024925) later undergoing two rounds of duplication events in lamprey and found to affect human cognitive abilities (Prosselkov et al., 2015). NTNG paralogs are expressed in the human neural crest-forming cells with NTNG2 10 times stronger than NTNG1 (Rada-Iglesias et al., 2012), both are differentially expressed in human comparing to chimpanzee and rhesus monkey with NTNG2 expression model showing stronger probability than NTNG1 (Iskow et al., 2012), and both are stronger expressed in human telencephalon comparing to chimpanzee and macaque (Konopka et al., 2012). NTNG1 has been classified
as a brain module hub gene “whose pattern fundamentally shifted between species” (Hawrylycs et al., 2015). Belonging to distinct modules of brain expression regulation (Liu et al., 2012, Konopka et al., 2012), NTNGs are classified as “genes with human-specific expression profiles” (Liu et al., 2012). The nearby gene ~260 kbp upstream of NTNG2 is MED27 (mediator of RNA polymerase II) has been proposed to be associated with the evolution of human-specific traits (McLean et al., 2011). NTNG1 has also been reported among the “adaptive plasticity genes” (Ghalambor et al., 2015) potentiating rapid adaptive evolution in guppies (NTNG2 was not found among the input RNA for analysis).

Complementarity among the NTNG paralogs and encoded by them proteins has been reported previously: brain expression complementary pattern (in almost self-exclusive manner) defined by the 5’-UTR-localised cis-regulatory elements (Yaguchi et al., 2014); complementary distribution within the hippocampal laminar structures (Nishimura-Akiyoshi et al., 2007); axon-dendrite synaptic ending resulting in differential control over the neuronal circuit plasticity (Matsukawa et al., 2014); mutually-exclusive binding pattern to postsynaptic partners, NGL-1 and NGL-2, dictated by the protein unstructural elements (Seiradake et al., 2011); alternative promoter usage vs alternative mRNA splicing (Wu et al., 2012) and increased coefficient of variation (CV, ST1d) for NTNG1 expression but not NTNG2 in SCZ patients (similar to Zhang et al., 2015); KO mice behavioral phenotypes and subcellular signaling partners complementarity (Zhang et al., 2016); “differential stability” brain modules expression (NTNG1 is expressed in the dorsal thalamus (M11) as a hub gene (Pearson’s 0.92) while NTNG2 is in neocortex and claustrum module (M6, Pearson’s 0.65)) (Hawrylycs et al., 2015); hypocretin neurons-specific expression of NTNG1 (but not NTNG2) as a sleep modulator (Yelin-Bekerman et al., 2015); top-down vs bottom-up information flows gating in mice and differential responsiveness to neuronal stimuli (Prosselkov et al., forthcoming); and human IQ-compiling cognitive domains...
complementation (Prosselkov et al., 2015). The current study reports the NTNGs complementarity association with the CDs (Figure 1A); mRNA splicing pattern complementary at the quantitative and qualitative levels via differential use of the middle-located exons (Figure 1B); brain complementary oscillatory expression over the human life span observed in the intensive cognitively loaded brain areas (Figure 2); AE of the paralogs-segregated unique non-coding elements (Figure 3A); complementary pattern of the protein orthologs (mice-to-human) protein sequence evolution. Such multi-level complementation is likely to reflect a shared evolutionary origin from a single gene in a primitive vertebrate organism 700 mln yrs ago and its subsequent functional segregation among the evolution-generated gene duplicates in jawless fish, such as lamprey.

Occupying independent but intercalating functional niches, NTNG1 and NTNG2 do not compensate but complement each other’s function forming a “functional complement” of genes. Half a billion yrs ago the doubled gene dosage led to the gradual SF and manifested in a function complementation within the cognitive domains, at least in human. We would like to coin such gene pair as a Cognitive Complement (CC).

CONCLUSION

The emerged functional redundancy, as an outcome of gene duplication, leads to function subdivision and its bifurcation among the gene paralogs resulting in the paralogs SF. A functional compensation is known to exist among the evolutionary unrelated genes but has not been reported among the gene paralogs, more frequently characterized by the function complementation. Gene paralogs structural identity (at both, gene and protein levels) does not provide a substrate for functional compensation but rather for complementation, perturbing “structure drives function” rule. A gene duplication event of a tunicate NTNG primordial gene and the subsequent process of its function specialisation (driven by the new ecological
niches appearance and evolution) among the gene duplicates made them to SF into distinct cognitive domains in a complementary manner forming a CC. In our forthcoming work we are to describe how Ntg mice genes function resembles that of human orthologs (Prosselkov et al., forthcoming).
MATERIALS AND METHODS

**Human brain NTNG transcriptome reconstruction.** Relates to Figure 1B and 1-C. The original source of the dataset was produced by (Wu et al., 2012: E-MATB-1030) and the downloaded .bam files used for the re-processing are listed in **ST1a**. All reconstructed transcripts are presented in **ST1d** standalone Excel file. Two samples were excluded from the analysis due to failed “per base sequence quality” measure, and zero expression level for NTNG1a and NTNG1int(9-10) otherwise consistently expressed throughout other samples (**ST1b**). SAMtools software was used for the SNPs calling from the available RNA-seq datasets (**ST1c**). For details refer to SM.

**Human brain expression profiling for NTNGs across the life span.** The original source of data was [www.brainspan.org](http://www.brainspan.org). All available samples were initially included into the analysis but two of them excluded at a later stage (MD for 12-13 pcw and mPFC for 16-19 pcw) due to high deviation (6-7 times) from the mean for other replicas. The mean expression values per each brain area as RPKM were plotted against the sampling age. Profiles classification was done visually considering the trend over the all plotted points as an average.

**NTNG1 (NTNG1m) and NTNG2 (NTNG2b) full-length mRNA transcripts assembly.** Relates to Figure 3B. Human NTNG1m brain transcript has been reported previously (Meerabux et al., 2005) and we have also confirmed its ortholog presence in the mice brain via full-length cloning (Prosselkov et al., unpublished). Since NCBI contains only its partial CDS (AY764265), we used the RNA-seq-generated exons (Figure 1B) to reconstruct its full-length and to generate an ORF of the encoded Netrin-G1m. Similarly, human NTNG2b was reconstructed from the RNA-seq dataset and from Ensemble as follows. Exon 5 sequence was deduced from ENST00000372179, other exons were from ENST00000467453 (no longer available on the current version of Ensemble) except for exon 6 deduced by running three independent alignments against the human genomic DNA with the mice 3’-intron (5-6),
exon 6, and 5'-intron (6-7) concomitantly confirmed by the generated full-length ORF for Netrin-G2b. The reconstructed protein was predicted to encode 587 amino acids, which is in a close proximity to the mice netrin-G2b ortholog of 589 residues (Prosselkov et al., forth.).

**Full-length gene structures of NTNG paralogs reconstruction.** Relates to Figure 3A. Both, the obtained above from the STG brain samples RNA-seq and the reconstructed full-lengths transcripts carrying all stably expressed exons were used to confirm the intron-exon junctions positioning for NTNG1 and NTNG2. Due to observed variability in the intron (1-2) and exon 10 sizes their boundaries were left unmarked.

**SUPPLEMENTARY MATERIALS (SM)**

Contain Supplementary Methods (RNA-seq of STG re-processing and SNPs detection) and Supplementary Tables (ST1a-d, ST2) as a single compiled pdf file. Reconstructed RNA-seq (.gtf) of the STG is presented as a standalone Excel file (ST1d). Also included: Netrin-G2b predicted phosphorylation and O-glycosylation, Netrin-G1 vs Netrin-G2 Ukd alignment (McWilliam et al., 2013), predicted secretory peptide cleavage and GPI attachments.

**ACKNOWLEDGEMENTS**

This work was in part supported by the “Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)” initiated by the Council for Science and Technology Policy (CSTP), and KAKENHI 15H04290 from the Japan Society for the Promotion of Science (JSPS).

**COMPETING INTERESTS**

Authors would like to express a lack of any competing interests associated with the work.
REFERENCES


Ghalambor CK, Hoke KL, Ruell EW, Fischer EK, Reznick DN, Hughes KA. 2015. Non-


Guilloux JP, Gaiteri C, Sibille E. 2010. Network analysis of positional candidate genes of schizophrenia highlights ... more than ... myelin-related pathways. Molecular Psychiatry **15**:786-788. doi: 10.1038/mp.2009.68.


McInnis MG, McIntosh A, McKinney R, McLean AW, McMahon FJ, McMahon WM,
McQuillin A, Medeiros H, Medland SE, Meier S, Melle I, Meng F, Meyer J, Middeldorp
CM, Middleton L, Milanova V, Miranda A, Monaco AP, Montgomery GW, Moran JL,
Moreno-De-Luca D, Morken G, Morris DW, Morrow EM, Moskvina V, Muglia P,
Mühleisen TW, Muir WJ, Muller-Myhsok B, Murtha M, Myers RM, Myin-Germeys I,
Neale MC, Nelson SF, Nievergelt CM, Nikolov I, Nimgaonkar V, Nolen WA, Nöthen
MM, Nurnberger JI, Nwulia EA, Nyholt DR, O'Dushlaine C, Oades RD, Olincy A,
Oliveira G, Olsen L, Ophoff RA, Osby U, Owen MJ, Palotie A, Parr JR, Paterson AD,
Pato CN, Pato MT, Penninx BW, Pergadia ML, Pericak-Vance MA, Pickard BS, Pimm J,
Piven J, Posthumus D, Potash JB, Pousta F, Propping P, Puri V, Quested DJ, Quinn EM,
Ramos-Quiroga JA, Rasmussen HB, Raychaudhuri S, Rehnström K, Reif A, Ribasés M,
Rice JP, Rietschel M, Roeder K, Roeyers H, Rossin L, Rothenberger A, Rouleau G,
Ruderfer D, Rujescu D, Sanders AR, Sanders SJ, Santangelo SL, Sergeant JA, Schachar R,
Schalling M, Schatzberg AF, Scheftner WA, Schellenberg GD, Scherer SW, Schork NJ,
Schulze TG, Schumacher J, Schwarz M, Scolnick E, Scott LJ, Shi JX, Shilling PD, Shyn
SI, Silverman JM, Slager SL, Smalley SL, Smit JH, Smith EN, Sonuga-Barke EJS, St
Clair D, State M, Steffens M, Steinhausen HC, Strauss JS, Strohmaier J, Stroup TS,
Sutcliffe JS, Szatmari P, Szelinger S, Thirumalai S, Thompson RC, Todorov AA, Tozzi F,
Treatlein J, Uhr M, van den Oord EJCG, Van Grootheest G, Van Os J, Vicente AM,
International Inflammatory Bowel Disease Genetics Consortium (IIBDGC), Devlin B,
Keilson JR, Sklar P, Daly MJ, O'Donovan MC, Craddock N, Sullivan PF, Smoller JW,
Kendler KS, Wray NR. 2013. Genetic relationship between five psychiatric disorders


NTNG1 reported associations | Cognitive disorder (CD) | A-2 NTNG2 reported associations
--- | --- | ---
Voineagu et al., 2011; O’Roak et al., 2012a,b; King et al., 2013*; Iossifov et al., 2014; Sanders et al., 2015; D’Gama et al., 2015 | Autism sporadic disorder (ASD) | Sanders et al., 2015; D’Gama et al., 2015
Eastwood and Harrison, 2008, 2010; Akula et al., 2014 | Bipolar disorder (BD) | Eastwood and Harrison, 2008, 2010; Akula et al., 2014
Fukasawa et al., 2004; Aoki-Suzuki et al., 2005; JSSLG et al. 2005; Eastwood and Harrison, 2008; Ohtsuki et al., 2008; Zakharyan et al., 2011; Zhu et al., 2011; Ayalew et al., 2012; Wilcox and Quadri, 2014 | Schizophrenia (SCZ) | Aoki-Suzuki et al., 2005; Eastwood and Harrison, 2008
Zhang et al., 2013 | Alzheimer disease (AD) |
Lagier-Tourenne et al., 2012*; Ishigaki et al., 2012*; Rogelj et al., 2012*; Nakaya et al., 2013* | Amyotrophic lateral sclerosis (ALS) |
Fronto-temporal lobe dementia (FTLD) |
Maurano et al., 2012; Wang et al., 2011; Boraska et al., 2014 | Anorexia nervosa (AN) |
van Kuilenburg et al., 2009 | Dihydropyrimidine dehydrogenase deficiency (DPD) |
Gilissen et al., 2014 | Intellectual disability (ID) |
Stepanyan et al., 2013 | Ischemic stroke (IS) |
Bisgaard et al., 2007 | Mental retardation (MeR) |
Stewart et al., 2013* | Obsessive-compulsive disorder (OCD) |
Lesnick et al., 2007 | Parkinson disease (PD) |
Cavard et al., 2009 | Pseudopapillary neoplasm (PN) |
Borg et al., 2005; Archer et al., 2006; Nectoux et al., 2007 | Rett syndrome (RS) |
Systemic lupus erythematosus (SLE) |
Mauroano et al., 2012 | Temporal lobe epilepsy (TLE) | Pan et al., 2010
Scharf et al., 2013* | Tourette syndrome (TS) |

Figure 1. NTNG paralogs complementation within neurological disorders and brain transcriptome. (A-1, A-2) Reported cognitive disorders associations for NTNG1 and NTNG2. *denotes rather an indirect association via a direct interaction with the research target. (B-1, B-2) RNA-seq of the STG of healthy (circle) and SCZ (cross) human subjects. The original dataset was produced by (Wu et al., 2012), accession number E-MTAB-1030 on ArrayExpress (ST1a) and reprocessed as described in SM. Five NTNG1 and four NTNG2 transcripts, consistently expressed across all 16 human samples are shown. Two samples (one healthy and one SCZ) have been omitted due to unsatisfactory quality of reads and expression profiling (ST1b). For the SNPs calling by SAMtools see ST1c. Data are presented as a mean RPKM±SEM. (C) Total number of the assembled transcripts across all samples for both paralogs (see ST1d for the completely reconstructed transcriptome). Dash-outlined are co-spliced exon clusters.
Figure 2. *NTNG* paralogs expression dynamics classification (A-C) in the human brain across the life span. A (1-4): further subdivision of the classifier. RNA-seq data are from the BrainSpan ([www.brainspan.org](http://www.brainspan.org)) presented as a mean±SEM. TCx = temporary neocortex; OCx = occipital neocortex; PCx = parietal neocortex; LGE = lateral ganglionic eminence; CB(C) = cerebellum(cortex); DTH = dorsal thalamus; URL = upper (rostral) rombic limb; pcw = post-conception week. Two data points (MD, 12-13 pcw, and mPFC, 16-19 pcw) for *NTNG1* expression were omitted as they were 6-7 times different from the mean of other replicas. All processed brain samples are listed in ST2.
Figure 3. Human NTNG paralogs DNA and protein sequence comparisons and “structure-function” rule incongruency. (A) Identical gene structures with different sizes of introns. RNA-seq data from Figure 1B were used to precisely deduce the exon/intron junction boundaries. The sizes of exons 1, 10 and introns (1-2) are not indicated due to observed among the splice transcripts lengths variability (see ST1a for details). Arrows indicate location of CNS = conserved non-coding sequences underwent accelerated evolution in humans compare to mice (mCNS) and chimpanzee (chCNS), as per Prabhakar et al., 2006; and ASC = anthropoid-specific constrained regions in human compare to marmoset (maASC), as per del Rosario et al., 2014. (B) Identical exon composition of the longest NTNG encoded RNA paralog transcripts and corresponding proteins with relatively high percent of identity among them dependent on the included/excluded Ukd domain (B-2) encoded by the exons 6 and 7 (B-1). Notably, the protein sequence represents higher percent of the paralogs difference than encoded it DNA. The matrices were obtained by GeneJockey II (Biosoft). (C) Protein alignments for the longest human encoded proteins, Netrin-G1m and Netrin-G2b, with Loops I-III highlighting binding sites for their cognate post-synaptic binding partners NGL-1. (D) Identical structural motif of the Netrin-G1/NTNG1 and Netrin-G2/NTNG2 complexes as per Seiradake et al. (2011). The figure’s reproduction is covered by the Creative Commons license.