



# All-or-(N)One – an epistemological characterization of the human tumorigenic neuronal paralogous FAM72 gene loci



Arne Kutzner<sup>a</sup>, Subrata Pramanik<sup>b</sup>, Pok-Son Kim<sup>c</sup>, Klaus Heese<sup>b,\*</sup>

<sup>a</sup> Department of Information Systems, College of Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 133–791, Republic of Korea

<sup>b</sup> Graduate School of Biomedical Science and Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 133–791, Republic of Korea

<sup>c</sup> Department of Mathematics, Kookmin University, 77 Jeongneung-ro, Seongbuk-gu, Seoul 136–702, Republic of Korea

## ARTICLE INFO

### Article history:

Received 25 March 2015

Received in revised form 26 May 2015

Accepted 13 July 2015

Available online 20 July 2015

### Keywords:

FAM72  
neuron  
cancer  
homo  
p17

## ABSTRACT

FAM72 is a novel neuronal progenitor cell (NPC) self-renewal supporting protein expressed under physiological conditions at low levels in other tissues. Accumulating data indicate the potential pivotal tumorigenic effects of FAM72. Our *in silico* human genome-wide analysis (GWA) revealed that the FAM72 gene family consists of four human-specific paralogous members, all of which are located on chromosome (chr) 1. Unique asymmetric FAM72 segmental gene duplications are most likely to have occurred in conjunction with the paired genomic neighbour SRGAP2 (SLIT-ROBO Rho GTPase activating protein), as both genes have four paralogues in humans but only one vertebra-emerging orthologue in all other species. No species with two or three FAM72/SRGAP2 gene pairs could be identified, and the four exclusively human-defining orthologues, with different mutation patterns in *Homo neanderthalensis* and *Denisova hominin*, may remain under epigenetic control through long non-coding (lnc) RNAs.

© 2015 Published by Elsevier Inc.

## 1. Introduction

For decades, evolutionary biologists have used modern molecular biology to perform phylogenetic analyses to characterize species-specific gene activities [1–4]. The study of the human genome and its position on the phylogenetic tree of life can help identify genes responsible for higher brain functions. Further, as we learn more about the human genome, we are discovering protooncogenes involved in cancer [5,6]. It is particularly important to focus on understanding the specific genetic mutations and epigenetic influences in different types of cancer because the overall incidence of cancer is increasing rapidly, with a combined annual increase of approximately 3.5% across all cancers [7].

Recently, neuron-specific SRGAP2 was identified as a gene involved in dendritic spine formation [8], and neuronal BDNF and MAP2 were found not only to contribute to dendrite-dependent brain plasticity [9–11] but also to play pivotal roles in non-neuronal cancer [12,13]. The recently identified FAM72 protein (otherwise known as Ugene, LMPIP or p17) has been shown to be involved in brain neuronal signaling pathways [14] and NPC maintenance [15] and has demonstrated promising clinical relevance for the survival and outcome of patients with various types of non-neuronal cancers [14,16–18].

Here, we investigated the genomic representation of FAM72 to improve predictions regarding its involvement in higher brain functions and as a novel potential therapeutic target for the treatment of cancers. We report for the first time a unique set of two genes FAM72 and SRGAP2 that characterize the emergence of vertebrates and define the human species based on its cognitive functions.

## 2. Methods

### 2.1. Data sets

The human chromosome (chr) dataset was downloaded from the public database of the National Center for Biotechnology Information (NCBI). The detailed dataset for chr 1 has the access No. NC\_000001.11 (chr 1, GRCh38). The FAM72-related gene IDs and other data sources used in this study are provided as Supplementary data (Supplementary methods and Supplementary Fig. S1).

### 2.2. Data analysis

For the Neanderthal and the Denisova genomes, we extracted the FASTQ reads from the provided BAM files and realigned them locally to the GRCh38 assembly with the Bowtie2 aligner [19] with the standard parameters unchanged (Supplementary methods). For global multi-alignments, we relied on Muscle (Multiple Sequence Comparison by Log-Expectation) [20], and all standard parameters were unchanged unless stated otherwise.

\* Corresponding author.

E-mail addresses: [klaus@hanyang.ac.kr](mailto:klaus@hanyang.ac.kr), [klaus.heese@rub.de](mailto:klaus.heese@rub.de) (K. Heese).

### 2.3. Taxonomy analysis

We developed a special tailored C++11 application with an embedded Burrows-Wheeler Aligner [21] as the central component. Using the integrated aligner, our application allowed for the tracing of a given sequence within a taxonomic context. We relied on the taxonomy offered by NCBI and reconstructed the taxonomic tree from publically available database dumps. Additionally, we retrieved the genomes of species within the taxonomy from assemblies offered by NCBI, and all data retrieval occurred automatically by evaluating database-information available on NCBI-servers.

### 2.4. Cancer analysis

We analysed publically available cancer genome datasets at the cBioPortal for Cancer Genomics (Supplementary data) providing access to data from 20,958 tumour samples from 89 cancer studies (data available up to 21st January 2015) [22,23] and data from the 1000 Genome project [24–26] to identify mutations, copy-number alterations and mRNA expression levels (using a mRNA expression z-score threshold value of  $\pm 2.0$ ).

## 3. Results

### 3.1. Chr 1 contains four distinct loci for FAM72

Analysis of the entire human genome for FAM72 using a small nt homology search revealed that only chr 1 contains FAM72 gene family members. However, chr 1 contains four different loci (A–D) – one on the forward (+) and three on the reverse (–) strand. Remarkably, all four paralogous FAM72 genes have long introns (Supplementary Fig. S2, Figs. 1 and 2). Thus far, the latest GRCh38 predicts additional exons for FAM72 A and B, which in turn use alternate start codons to produce shorter isoforms. Considering prior information based on GRCh37, the existence of other *in vivo* isoforms is conceivable (Supplementary Fig. S2).

### 3.2. FAM72 short transcripts

When investigating the short FAM72 transcripts, NCBI's new GRCh38 gene assembly predicted a novel 363-bp (120 amino acids (aa)) gene transcript with an additional exon at the 5' site using the second in-frame ATG codon of its full-length counterpart. This observation is in contrast to the GRCh37.p13 gene assembly, which also predicted a short FAM72 transcript (330 bp, 109 aa) but used the same start codon as the long 450-bp (149 aa) variant with an alternative splicing pattern (Supplementary Figs. S2B and C). The newly predicted 363-bp (120 aa) transcript (using NCBI's Gnomon software [27]), however, remains questionable because the in-frame start codon has a rather weak Kozak sequence compared with the start codon used by the longer 450-bp and the older 330-bp transcripts (Supplementary Fig. S2C).

### 3.3. The four FAM72 CDSs on chr 1 differ at six positions

A comparison of the four different CDSs (450 nts) of FAM72 at the four different loci (A–D) on chr 1 revealed that they differ at six positions. The start codon area is embedded in a typical Kozak sequence [28], and the first 18 nts specifically identify FAM72 throughout the entire human genome (Supplementary Figs. S2C and D).

### 3.4. Five differences in the human FAM72 paralogous proteins

A comparison of the four FAM72 proteins encoded by the four different loci (A–D) on chr 1 also revealed that, based on the six nt differences

identified in the four different CDSs, only five differences result in nonsynonymous changes at the protein level. The variation observed at the nt 6 position of the CDS does not produce a difference at the protein level (Supplementary Figs. S2B and E).

### 3.5. FAM72's 5'-UTRs and 3'-UTRs

FAM72 5'-UTRs exhibit differences at only five positions, and member A has a shorter known 5'-UTR. A comparison of all the 3'-UTRs indicated that they are highly homologous, with only a few variations at nine different positions, and are typically AT-rich with potential regulatory regions (Supplementary Figs. S2F and G) [29]. The FAM72 mRNA matches the 2-kbp size average (5'-UTR + CDS + 3'-UTR, excluding poly(A)) but is rather non-standard with respect to its 3'-/5'-UTR ratio (Supplementary Table S1), with ratios ranging from 1.18 (FAM72C) to 2.60 (FAM72A long transcript) [30].

### 3.6. FAM72's GC and AT content

Because all FAM72 members are highly homologous, the overview of FAM72 GC and AT contents (Supplementary Table S2) shows that their GC and AT contents are almost identical except for the 5'-UTR of member A (GC content smaller than AT content), which is shorter than all the other 5'-UTRs. The GC content is higher than the AT content only in the 5'-UTR sequence. The 3'-UTR AT content of approximately 68% is higher than that of any other segment of the entire sequence. The FAM72 proteins are encoded by a higher AT content (approximately 58%) than GC content, and the AT content is particularly high in the last exon 4. In addition, the GC content at position 3 (GC3) of the CDS is rather low (approximately 47%).

### 3.7. Common gene family neighbours of FAM72 loci

By comparing the four different gene loci of human FAM72, we discovered that they all resided next to the same four paralogous genes: SRGAP2 A–D (Fig. 1A) while all these four gene pairs are separated by homologous long non-coding (lnc) RNAs. Our comparison revealed that all FAM72 members (A–D) had uniform intron lengths, whereas the intron lengths of SRGAP2 members varied (Supplementary Figs. S2A, B and Supplementary Table S3).

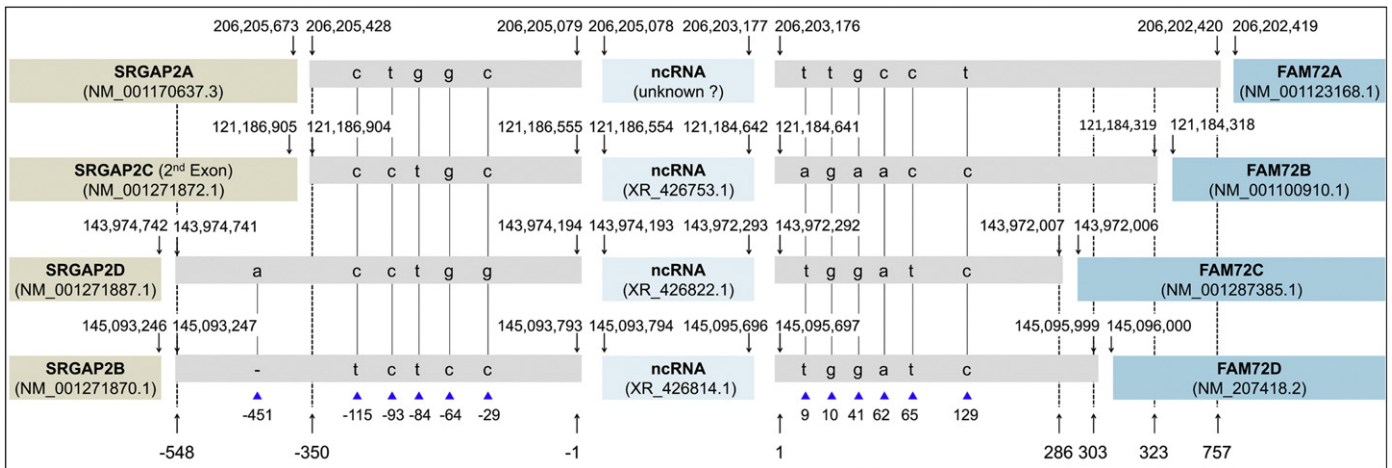
A closer examination of the interspace between SRGAP2 and FAM72 indicated that FAM72 is a stable constant for this duplicated genome area, whereas the interspace displays rather complex variation, with the appearance of a new gene (lncRNA) with an as yet undefined exon structure (Figs. 1B, C and 2).

### 3.8. FAM72's orthologues

Interestingly, both gene families, FAM72 and SRGAP2, were composed of four human paralogues located as unique pairs on chr 1 (Fig. 1A). However, all other species investigated contained only one orthologue (Fig. 3) [31]. Thus, FAM72 and SRGAP2 appear to represent a unique gene couple that characterizes the emergence of vertebrates with a notochord (with one gene pair) and defines the human species containing four gene pairs (Fig. 3). The earliest emergence of the FAM72A/SRGAP2A gene pair has been identified in *Danio rerio* (Fig. 3).

Comparing the phylogenetic taxonomy of all individual FAM72 exons indicated that exons 1–3 are more stable than exon 4 and that primates are more closely related to Laurasiatheria (including cat, dog, and horse) than to Murinae (mouse and rat). The comparison also revealed two stable aa regions, on exons 1 and 3, and five variable aa regions that were observed in primates and Laurasiatheria but not in Murinae, indicating the importance of these regions for higher brain functions (Figs. 4 and 5).



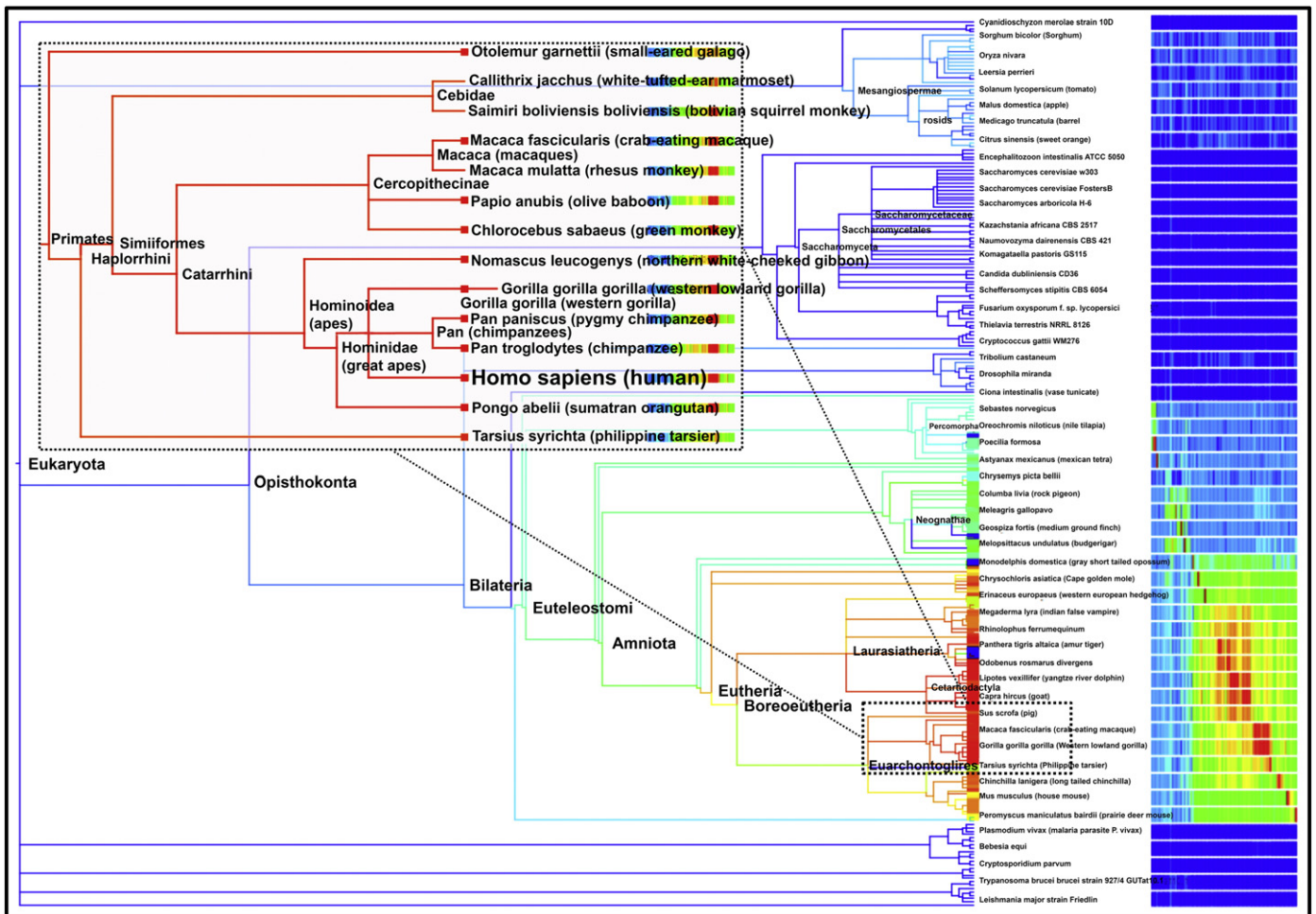


**Fig. 2.** Comparison of DNA intervals between FAM72 and SRGAP2 gene loci on chr 1. The nts from +287 to +757 between the lncRNAs and the FAM72A-D sequences are highly similar, where some part of an intron belongs to the first exon of another homologous FAM72 (see also Supplementary Fig. S2F). The nts from -351 to -548 between the lncRNAs and the SRGAP2A-D sequences are also highly similar, where some part of an intron belongs to the first exon of another homologous SRGAP2s (see also Supplementary Fig. S3).

If human, mouse and cat had a common ancestor, it could most likely be explained by the changes that occurred at aa 52, 111, 131 and 138. However, the changes in aa 147 suggest a common ancestor for *Euarchontoglires* (primates and *Murinae*) and *Laurasiatheria*. At this point, either *Murinae* degenerated or *primates* and *Laurasiatheria* together diverged from *Murinae* (Fig. 5).

### 3.9. FAM72 and Homo

The *Hominidae* form a taxonomic family of primates, including four extant genera: chimpanzees, gorillas, humans, and orangutans [33–35]. Recent genome analysis has demonstrated that Denisovans were cousins of the Neanderthals and interbred probably with *Homo erectus* or



**Fig. 3.** FAM72 in the phylogenetic tree. A search for all four FAM72 exons across all species. Red = high homology, blue = low homology. All four exons were identified in vertebrates and appear as a pair with SRGAP2 as early as in *Danio rerio*. *Drosophila* does not have FAM72. Species analysis for SRGAP2 can be found in Supplementary Fig. S4.



Supplementary Fig. S5. Previous cBioportal analyses have also indicated FAM72B (among others) as a potential biomarker of prostate cancer [38].

#### 4. Discussion

We unravelled that the four human paralogous genes FAM72A–D have the four paralogous genes SRGAP2A–D as their paired neighbours. Most likely, the lncRNAs between the FAM72/SRGAP2 gene pairs control tissue-specific expression patterns [39,40]. We could not confirm an unusual intron length compared with ten other randomly selected genes (data not shown) [27]. However, the low GC3 content may explain the low expression levels of the FAM72 proteins [41–43].

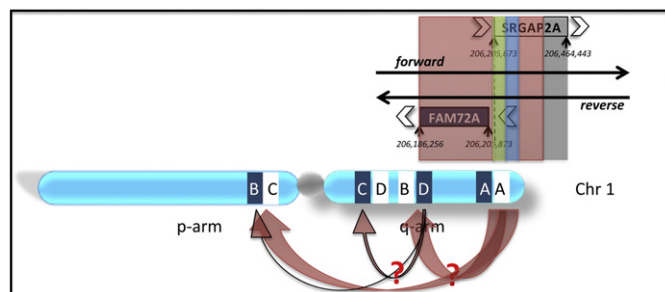
Human brain cognitive functions, including memory and learning processes, greatly depend on synaptic plasticity, which is largely controlled by exon–dendrite interactions [44–46]. SRGAP2 has been associated with neurodevelopmental disorders [47] and, similar to FAM72, the SRGAP2 paralogues display a broad tissue expression pattern with higher levels in neuronal tissue, particularly in the human brain, where they regulate neurodevelopmental processes [31,48–51]. Recently, SRGAP2 was demonstrated to have paralogous gene-specific functions during brain development. Dysfunction of a SRGAP2 paralogue may result in severe brain diseases, ranging from epileptic seizures to cognitive deficits [48]. Duplication of human-specific genes is particularly important for genes involved in nervous system development [52,53] but is also a genetic mechanism for evolutionary innovation in organismal adaptation [54].

WGD is associated with both brain disease (uncontrolled gene duplication) [55–60] and the origin of the human species [61]. However, genetic evolution for the innovative brain development required for higher brain functions in *Homo sapiens* and genetic variation-based brain diseases have been considered to be closely related [62,63].

From an evolutionary point of view, the FAM72/SRGAP2 gene pair seems to define the notochord-containing vertebrates, whereas the four paralogous FAM72/SRGAP2 couples seem to be a distinctive feature of humans. Because they are observed exclusively in humans, these four FAM72/SRGAP2 gene couples define the emergence of *Homo sapiens* as a distinct species of *Hominidae* because of the peculiar developmental feature that defines the human brain with its higher cognitive functions [53]. Assuming that the gene duplication was completed with the appearance of *Homo australopithecus* [31], it remains an enigma why no known species contains two or three FAM72 (and SRGAP2) genes – the other great apes (e.g., chimpanzee and gorilla) contain only one such gene pair. Notably, although all four FAM72 genes in the genus of hominids (*Homo*) underwent the same (post-duplication) phylogenetic development, they have distinct mutation patterns, raising questions regarding current gene evolution theories.

A simple evolution-based asymmetric segmental gene duplication, which generated four human ohnologues [64] or positional paralogues (topparalogues) [65], as hypothetically previously discussed [53], seems to be epistemologically questionable because highly conserved areas with uniform exons and introns of both FAM72 and SRGAP2 genes are interrupted by variable parts of chr 1 (Figs. 2 and 8). Because the FAM72/SRGAP2 pair follows an ‘All (*Homo*) or (N)One’ (all other species) principle, another theory must be developed to explain this observation because the parameter time, as suggested previously [8,31], does not adequately explain this quantum leap between humans and the other great apes.

The identification of the NPC self-renewal supporting FAM72 gene [15] as a cancer-specific candidate is highly promising. Thus far, the phenomenon of (brain-) neuron-specific genes expressed in non-



**Fig. 8.** Comparison of the paired FAM72A and SRGAP2A gene loci on chr 1. Gene duplication (SRGAP2A → SRGAP2B followed by SRGAP2B → SRGAP2C and D) had previously been suggested. Non-preserved (green and blue) areas (see Figs. 1C, and 2) in the middle of preserved (red) areas suggest against FAM72 ohnology (WGD) by simple asymmetric segmental gene duplication (if not followed by a partial interregional SRGAP2 exon degeneration/deletion and intron variation, including lncRNA generation and variations). The grey area was lost in all truncated SRGAP2 versions B–D.

neuronal tissue has been rarely observed and remains poorly understood [12,66,67]. It is tempting to speculate that, under physiological conditions, FAM72 is active in the nervous system to maintain the self-renewal capacity of NPCs [8,14,15]. However, under pathophysiological conditions, FAM72 may cause post-mitotic neuronal cell death [14] and become a pivotal player in various cancer diseases in other tissues, particularly in breast cancer with higher PKC expression [16,17] (Fig. 4). Because FAM72 seems to characterize pathologically proliferating cells outside of the nervous system but is otherwise absent, it is an interesting therapeutic drug target candidate in cancer [68]. In this regard, it is notable that a GWA in East Asians identified a breast cancer susceptibility locus at 1q32.1 (FAM72A) within intron 2 of ZC3H11A and the 5' UTR of the ZBED6 (~2300 kbp downstream of FAM72A) [69]. By contrast, thus far, SRGAP2 has not been linked to any cancer disease, and the biological relevance of these conjoint paralogous gene loci (FAM72 and SRGAP2) remains to be elucidated.

Three of the A–D FAM72 aa variations are located near the potential transmembrane domain (aa 74 – aa 94) of FAM72 [17], and the major changes at aa positions 94 (L → P), 99 (G → R) and 125 (W → R) (Supplementary Fig. S2 [70,71]) may indicate opposing functions [48].

#### 5. Conclusion

The biological relevance of the FAM72 ohnology (neo-functionalization or degeneration because of the new genomic context and aa variations) and the extent to which each FAM72 paralogue contributes to its physiological or pathophysiological outcome remains to be investigated. The lncRNAs between the FAM72/SRGAP2 gene pairs may play a pivotal role in the control of these gene pairs to guide their specific contributions to neuronal development (FAM72 in NPCs) and neuronal activities (SRGAP2 in post-mitotic neurons) [39,40].

#### Conflict of interest

The authors declare no competing financial interests.

#### Acknowledgments

This study was supported by Hanyang University. We thank the people at the Max Planck Institute for Evolutionary Anthropology (eva.mpg.de) for providing the Neanderthal and Denisova genome data.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ygeno.2015.07.003>.

## References

- [1] K. Wakasugi, N. Takahashi, H. Uchida, S. Watanabe, Species-specific functional evolution of neuroglobin, *Mar. Genomics* 4 (2011) 137–142.
- [2] X. Xin, D. Gfeller, J. Cheng, R. Tonikian, L. Sun, A. Guo, L. Lopez, A. Pavlenco, A. Akintobi, Y. Zhang, J.F. Rual, B. Currell, S. Seshagiri, T. Hao, X. Yang, Y.A. Shen, K. Salehi-Ashtiani, J. Li, A.T. Cheng, D. Bouamalay, A. Lugari, D.E. Hill, M.L. Grimes, D.G. Drubin, B.D. Grant, M. Vidal, C. Boone, S.S. Sidhu, G.D. Bader, SH3 interactome conserves general function over specific form, *Mol. Syst. Biol.* 9 (2013) 652.
- [3] P. Rakic, Evolution of the neocortex: a perspective from developmental biology, *Nat. Rev. Neurosci.* 10 (2009) 724–735.
- [4] H. Baker, Evaluation of species-specific biochemical variation as a means for assessing homology in neuronal populations, *Brain Behav. Evol.* 38 (1991) 255–263.
- [5] C. Napoli, M. Sessa, T. Infante, A. Casamassimi, Unraveling framework of the ancestral Mediator complex in human diseases, *Biochimie* 94 (2012) 579–587.
- [6] R.T. Moon, S.T.K.E. Sci, Wnt/beta-catenin pathway, 2005 (2005) cm1.
- [7] K.W. Jung, Y.J. Won, H.J. Kong, C.M. Oh, H.G. Seo, J.S. Lee, Cancer statistics in Korea: incidence, mortality, survival and prevalence in 2010, *Cancer Res. Treat.* 45 (2013) 1–14.
- [8] D.H. Geschwind, G. Konopka, Neuroscience: Genes and human brain evolution, *Nature* 486 (2012) 481–482.
- [9] N. Kuczewski, C. Porcher, V. Lessmann, I. Medina, J.L. Gaiarsa, Activity-dependent dendritic release of BDNF and biological consequences, *Mol. Neurobiol.* 39 (2009) 37–49.
- [10] L.A. Roberts, C.H. Large, M.J. Higgins, T.W. Stone, C.T. O'Shaughnessy, B.J. Morris, Increased expression of dendritic mRNA following the induction of long-term potentiation, *Brain Res. Mol. Brain Res.* 56 (1998) 38–44.
- [11] J. Zhong, T. Zhang, L.M. Bloch, Dendritic mRNAs encode diversified functionalities in hippocampal pyramidal neurons, *BMC Neurosci.* 7 (2006) 17.
- [12] M.H. Soltani, R. Pichardo, Z. Song, N. Sangha, F. Camacho, K. Satyamoorthy, O.P. Sanguera, V. Setaluri, Microtubule-associated protein 2, a marker of neuronal differentiation, induces mitotic defects, inhibits growth of melanoma cells, and predicts metastatic potential of cutaneous melanoma, *Am. J. Pathol.* 166 (2005) 1841–1850.
- [13] L.A. Liotta, E. Kohn, Anokis: cancer and the homeless cell, *Nature* 430 (2004) 973–974.
- [14] S. Nehar, M. Mishra, K. Heese, Identification and characterisation of the novel amyloid-beta peptide-induced protein p17, *FEBS Lett.* 583 (2009) 3247–3253.
- [15] B.A. Benayoun, E.A. Pollina, D. Ucar, S. Mahmoudi, K. Karra, E.D. Wong, K. Devarajan, A.C. Daugherty, A.B. Kundaje, E. Mancini, B.C. Hitz, R. Gupta, T.A. Rando, J.C. Baker, M.P. Snyder, J.M. Cherry, A. Brunet, H3K4me3 breadth is linked to cell identity and transcriptional consistency, *Cell* 158 (2014) 673–688.
- [16] C. Guo, X. Zhang, S.P. Fink, P. Platzer, K. Wilson, J.K. Willson, Z. Wang, S.D. Markowitz, Ugene, a newly identified protein that is commonly overexpressed in cancer and binds uracil DNA glycosylase, *Cancer Res.* 68 (2008) 6118–6126.
- [17] K. Heese, The protein p17 signaling pathways in cancer, *Tumour Biol.* 34 (2013) 4081–4087.
- [18] L.T. Wang, C.S. Lin, C.Y. Chai, K.Y. Liu, J.Y. Chen, S.H. Hsu, Functional interaction of Ugene and EBV infection mediates tumorigenic effects, *Oncogene* 30 (2011) 2921–2932.
- [19] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, *Nat. Methods* 9 (2012) 357–359.
- [20] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.* 32 (2004) 1792–1797.
- [21] H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform, *Bioinformatics* 26 (2010) 589–595.
- [22] J. Gao, B.A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S.O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, N. Schultz, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal, *Sci. Signal.* 6 (2013) p11.
- [23] E. Cerami, J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, B.A. Aksoy, A. Jacobsen, C.J. Byrne, M.L. Heuer, E. Larsson, Y. Antipin, B. Reva, A.P. Goldberg, C. Sander, N. Schultz, The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, *Cancer Discov.* 2 (2012) 401–404.
- [24] C. Genomes Project, G.R. Abecasis, A. Auton, L.D. Brooks, M.A. DePristo, R.M. Durbin, R.E. Handsaker, H.M. Kang, G.T. Marth, G.A. McVean, An integrated map of genetic variation from 1,092 human genomes, *Nature* 491 (2012) 56–65.
- [25] E. Khurana, Y. Fu, V. Colonna, X.J. Mu, H.M. Kang, T. Lappalainen, A. Sboner, L. Lozhovskiy, J. Chen, A. Harmanci, J. Das, A. Abyzov, S. Balasubramanian, K. Beal, D. Chakravarty, D. Challis, Y. Chen, D. Clarke, L. Clarke, F. Cunningham, U.S. Evani, P. Flicek, R. Fragoza, E. Garrison, R. Gibbs, Z.H. Gumus, J. Herrero, N. Kitabayashi, Y. Kong, K. Lage, V. Liliushvili, S.M. Lipkin, D.G. MacArthur, G. Marth, D. Muzny, T.H. Pers, G.R. Ritchie, J.A. Rosenfeld, C. Siu, X. Wei, M. Wilson, Y. Xue, F. Yu, C. Genomes Project, E.T. Dermitzakis, H. Yu, M.A. Rubin, C. Tyler-Smith, M. Gerstein, Integrative annotation of variants from 1092 humans: application to cancer genomics, *Science* 342 (2013) 1235587.
- [26] L. Clarke, X. Zheng-Bradley, R. Smith, E. Kulesha, C. Xiao, I. Toneva, B. Vaughan, D. Preuss, R. Leinonen, M. Shumway, S. Sherry, P. Flicek, C. Genomes Project, The 1000 Genomes Project: data management and community access, *Nat. Methods* 9 (2012) 459–462.
- [27] A. Souvorov, Y. Kapustin, B. Kiryutin, V. Chetvermin, T. Tatusova, D. Lipman, Gnomon – NCBI eukaryotic gene prediction tool, National Center for Biotechnology Information, 2010. 1–24.
- [28] M. Kozak, Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes, *Cell* 44 (1986) 283–292.
- [29] K. Heese, Y. Nagai, T. Sawada, The 3' untranslated region of the new rat synaptic vesicle protein 2B mRNA transcript inhibits translational efficiency, *Brain Res. Mol. Brain Res.* 104 (2002) 127–131.
- [30] W. Makalowski, M.S. Boguski, Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences, *Proc. Natl. Acad. Sci. U. S. A.* 95 (1998) 9407–9412.
- [31] M.Y. Dennis, X. Nuttle, P.H. Sudmant, F. Antonacci, T.A. Graves, M. Nefedov, J.A. Rosenfeld, S. Sajjadian, M. Malig, H. Kotkiewicz, C.J. Curry, S. Shafer, L.G. Shaffer, P.J. de Jong, R.K. Wilson, E.E. Eichler, Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication, *Cell* 149 (2012) 912–922.
- [32] S.H. White, Amino acid preferences of small proteins. Implications for protein stability and evolution, *J. Mol. Biol.* 227 (1992) 991–995.
- [33] K.E. Langergraber, K. Prufer, C. Rowley, C. Boesch, C. Crocford, K. Fawcett, E. Inoue, M. Inoue-Muruyama, J.C. Mitani, M.N. Muller, M.M. Robbins, G. Schubert, T.S. Stoinski, B. Viola, D. Watts, R.M. Wittig, R.W. Wrangham, K. Zuberbühler, S. Paabo, L. Vigilant, Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution, *Proc. Natl. Acad. Sci. U. S. A.* 109 (2012) 15716–15721.
- [34] Y. Gilad, M. Przeworski, D. Lancet, Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates, *PLoS Biol.* 2 (2004) E5.
- [35] S.E. Fisher, G.F. Marcus, The eloquent ape: genes, brains and the evolution of language, *Nat. Rev. Genet.* 7 (2006) 9–20.
- [36] M. Meyer, M. Kircher, M.T. Gansauge, H. Li, F. Racimo, S. Mallick, J.G. Schraiber, F. Jay, K. Prufer, C. de Filippo, P.H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R.E. Green, K. Bryc, A.W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M.F. Hammer, M.V. Shunkov, A.P. Derevianko, N. Patterson, A.M. Andres, E.E. Eichler, M. Slatkin, D. Reich, J. Kelso, S. Paabo, A high-coverage genome sequence from an archaic Denisovan individual, *Science* 338 (2012) 222–226.
- [37] M. Marshall, Mystery human species emerges from Denisovan genome, *New Sci.* (2013) (19 November 2013).
- [38] P. Rajan, J. Stockley, I.M. Sudbery, J.T. Fleming, A. Hedley, G. Kalna, D. Sims, C.P. Ponting, A. Heger, C.N. Robson, R.M. McMenamin, I.D. Pedley, H.Y. Leung, Identification of a candidate prognostic gene signature by transcriptome analysis of matched pre- and post-treatment prostatic biopsies from patients with advanced prostate cancer, *BMC Cancer* 14 (2014) 977.
- [39] T. Derrien, R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D.G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J.B. Brown, L. Lipovich, J.M. Gonzalez, M. Thomas, C.A. Davis, R. Shiekhattar, T.R. Gingeras, T.J. Hubbard, C. Notredame, J. Harrow, R. Guigo, The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression, *Genome Res.* 22 (2012) 1775–1789.
- [40] J.L. Rinn, H.Y. Chang, Genome regulation by long noncoding RNAs, *Annu. Rev. Biochem.* 81 (2012) 145–166.
- [41] S. Arhondakis, O. Clay, G. Bernardi, GC level and expression of human coding sequences, *Biochem. Biophys. Res. Commun.* 367 (2008) 542–545.
- [42] G. Kudla, L. Lipsinki, F. Caffin, A. Helwak, M. Zyllicz, High guanine and cytosine content increases mRNA levels in mammalian cells, *PLoS Biol.* 4 (2006) e180.
- [43] R. Robinson, More GC means more RNA, *PLoS Biol.* 4 (2006) e206.
- [44] C. Cunha, R. Brambilla, K.L. Thomas, A simple role for BDNF in learning and memory? *Front. Mol. Neurosci.* 3 (2010) 1.
- [45] J.L. Rekart, M.R. Holahan, A. Routtenberg, Presynaptic Structural Plasticity and Long-Term Memory: Focus on Learning-Induced Redistribution of Hippocampal Mossy Fibers, in: F. Bermudez-Rattoni (Ed.), *Neural Plasticity and Memory: From Genes to Brain Imaging*, Boca Raton (FL), 2007.
- [46] V. Ramirez-Amaya, Molecular Mechanisms of Synaptic Plasticity Underlying Long-Term Memory Formation, in: F. Bermudez-Rattoni (Ed.), *Neural Plasticity and Memory: From Genes to Brain Imaging*, Boca Raton (FL), 2007.
- [47] E.E. Govek, S.E. Newey, L. Van Aelst, The role of the Rho GTPases in neuronal development, *Genes Dev.* 19 (2005) 1–49.
- [48] C. Charrier, K. Joshi, J. Coutinho-Budd, J.E. Kim, N. Lambert, J. de Marchena, W.L. Jin, P. Vanderhaeghen, A. Ghosh, T. Sassa, F. Polleux, Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation, *Cell* 149 (2012) 923–935.
- [49] S. Guerrier, J. Coutinho-Budd, T. Sassa, A. Gresset, N.V. Jordan, K. Chen, W.L. Jin, A. Frost, F. Polleux, The F-BAR domain of srGAP2 induces membrane protrusions required for neuronal migration and morphogenesis, *Cell* 138 (2009) 990–1004.
- [50] S. Guo, S. Bao, srGAP2 arginine methylation regulates cell migration and cell spreading through promoting dimerization, *J. Biol. Chem.* 285 (2010) 35133–35141.
- [51] H. Saitsu, H. Osaka, S. Sugiyama, K. Kurosawa, T. Mizuguchi, K. Nishiyama, A. Nishimura, Y. Tsurusaki, H. Doi, N. Miyake, N. Harada, M. Kato, N. Matsumoto, Early infantile epileptic encephalopathy associated with the disrupted gene encoding Slit-Robo Rho GTPase activating protein 2 (SRGAP2), *Am. J. Med. Genet. A* 158A (2012) 199–205.
- [52] P.H. Sudmant, J.O. Kitzman, F. Antonacci, C. Alkan, M. Malig, A. Tsalenko, N. Sampas, L. Bruhn, J. Shendure, E.E. Eichler, Diversity of human copy number variation and multicopy genes, *Science* 330 (2010) 641–646.
- [53] D.H. Geschwind, P. Rakic, Cortical evolution: judge the brain by its cover, *Neuron* 80 (2013) 633–647.
- [54] W. Qian, J. Zhang, Genomic evidence for adaptation by gene duplication, *Genome Res* 24 (2014) 1356–1362.
- [55] M.F. Portnoi, Microduplication 22q11.2: a new chromosomal syndrome, *Eur. J. Med. Genet.* 52 (2009) 88–93.
- [56] T. Mattina, O. Palumbo, R. Stallone, R.M. Pulvirenti, L. Di Dio, P. Pavone, M. Carella, L. Pavone, Interstitial 16p13.3 microduplication: case report and critical review of genotype-phenotype correlation, *Eur. J. Med. Genet.* 55 (2012) 747–752.
- [57] S. Kitsiou-Tzeli, M. Tzetzis, C. Sofocleous, C. Vrettou, A. Xaidara, K. Giannikou, A. Pampanos, A. Mavrou, E. Kanavakis, De novo interstitial duplication of the 15q11.2-q14 PWS/AS region of maternal origin: Clinical description, array CGH analysis, and review of the literature, *Am. J. Med. Genet. A* 152A (2010) 1925–1932.
- [58] N. Hanemaaijer, T. Dijkhuizen, M. Haadsma, M. Boeve, M. Boon, R. Hordijk, K. Kok, B. Sikkema-Raddatz, C.M. van Ravenswaaij-Arts, A 649 kb microduplication in 1p34.1,

- including POMGNT1, in a patient with microcephaly, coloboma and laryngomalacia; and a review of the literature, *Eur. J. Med. Genet.* 52 (2009) 116–119.
- [59] L. Rodriguez, S.S. Bhatt, M. Garcia-Castro, A. Plasencia, J. Fernandez-Toral, E. Abarca, M. de Bello Cioffi, T. Liehr, A unique case of a discontinuous duplication 3q26.1-3q28 resulting from a segregation error of a maternal complex chromosomal rearrangement involving an insertion and an inversion, *Gene* 535 (2014) 165–169.
- [60] Z. Li, J. Liu, H. Li, Y. Peng, W. Lv, Z. Long, D. Liang, L. Wu, Phenotypic expansion of the interstitial 16p13.3 duplication: a case report and review of the literature, *Gene* 531 (2013) 502–505.
- [61] T.J. Crow, The 'big bang' theory of the origin of psychosis and the faculty of language, *Schizophr. Res.* 102 (2008) 31–52.
- [62] T.J. Crow, Schizophrenia as variation in the sapiens-specific epigenetic instruction to the embryo, *Clin. Genet.* 81 (2012) 319–324.
- [63] M. Lam, S.L. Collinson, K. Sim, C.E. Mackay, A.C. James, T.J. Crow, Asymmetry of lexico-semantic processing in schizophrenia changes with disease progression, *Schizophr. Res.* 134 (2012) 125–130.
- [64] K. Wolfe, Robustness—it's not where you think it is, *Nat. Genet.* 25 (2000) 3–4.
- [65] C.N. Dewey, Positional orthology: putting genomic evolutionary relationships into context, *Brief. Bioinform.* 12 (2011) 401–412.
- [66] W. Chen, J. Heierhorst, J. Brosius, H. Tiedge, Expression of neural BC1 RNA: induction in murine tumours, *Eur. J. Cancer* 33 (1997) 288–292.
- [67] A. Iacoangeli, Y. Lin, E.J. Morley, I.A. Muslimov, R. Bianchi, J. Reilly, J. Weedon, R. Diallo, W. Bocker, H. Tiedge, BC200 RNA in invasive and preinvasive breast cancer, *Carcinogenesis* 25 (2004) 2125–2133.
- [68] S. Pramanik, A. Kutzner, K. Heese, Lead discovery and in silico 3D structure modeling of tumorigenic FAM72A (p17), *Tumour Biol.* 36 (2015) 239–249.
- [69] Q. Cai, B. Zhang, H. Sung, S.K. Low, S.S. Kweon, W. Lu, J. Shi, J. Long, W. Wen, J.Y. Choi, D.Y. Noh, C.Y. Shen, K. Matsuo, S.H. Teo, M.K. Kim, U.S. Khoo, M. Iwasaki, M. Hartman, A. Takahashi, K. Ashikawa, K. Matsuda, M.H. Shin, M.H. Park, Y. Zheng, Y.B. Xiang, B.T. Ji, S.K. Park, P.E. Wu, C.N. Hsiung, H. Ito, Y. Kasuga, P. Kang, S. Mariapun, S.H. Ahn, H.S. Kang, K.Y. Chan, E.P. Man, H. Iwata, S. Tsugane, H. Miao, J. Liao, Y. Nakamura, M. Kubo, D.G.-O. Consortium, R.J. Delahanty, Y. Zhang, B. Li, C. Li, Y.T. Gao, X.O. Shu, D. Kang, W. Zheng, *Nat. Genet.* 886–890, Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1, 46 (2014) 886–890.
- [70] B. Palanisamy, K. Heese, Oxygen distribution in proteins defines functional significance of the genome and proteome of the malaria parasite *Plasmodium falciparum* 3D7, *FEMS Microbiol. Lett.* 351 (2014) 59–63.
- [71] B. Palanisamy, R. Ekambaram, K. Heese, Thymine distribution in genes provides novel insight into the functional significance of the proteome of the malaria parasite *Plasmodium falciparum* 3D7, *Bioinformatics* 30 (2014) 597–600.