

# Confirmation of region-specific patterns of gene expression in the human brain

Carl Ernst · Adolfo Sequeira · Tim Klempan ·  
Neil Ernst · Jarlath ffrench-Mullen · Gustavo Turecki

Received: 7 November 2006 / Accepted: 16 February 2007 / Published online: 21 March 2007  
© Springer-Verlag 2007

**Abstract** The human brain is divided and categorized in different ways, yet a molecular genetic approach to region specificity does not exist. Using data from 12 healthy control subjects across 18 brain regions, we performed a microarray analysis using both the HG-U133AB and HG-U133 plus 2 chips for each subject to determine molecular targets showing region specificity. Using a previously published data as our guide, we confirm SIX3, GPR6, SH3RF2, and hSyn as molecular markers of the nucleus accumbens and gamma-aminobutyric-acid A receptor alpha-6, Nik-related kinase, and eomesodermin as molecular markers of the cerebellum.

**Keywords** Brain · Microarray · Genetic marker

**Electronic supplementary material** The online version of this article (doi:10.1007/s10048-007-0084-2) contains supplementary material, which is available to authorized users.

C. Ernst · A. Sequeira · T. Klempan · G. Turecki  
McGill Group for Suicide Studies, McGill University,  
Montreal, QC, Canada

N. Ernst  
Department of Computer Science, University of Toronto,  
Toronto, ON, Canada

J. ffrench-Mullen  
Gene Logic, Inc,  
Gaithersburg, MD, USA

G. Turecki (✉)  
Douglas Hospital Research Centre, Pavilion Frank B Common,  
Rm. F-3125, 6875 LaSalle Blvd.,  
Verdun, Montreal, QC H4H 1R3, Canada  
e-mail: gustavo.turecki@mcgill.ca

## Introduction

To date, our understanding of the human brain is limited; regions are defined by gross anatomy, external morphology, or the location of specific nuclei and neuronal pathways [1, 2]. Brain region identification has progressed further with the introduction of antibodies that have led to the labeling of specific proteins in specific brain areas [3]. Still, given the complexity of the brain, much work needs to be done to understand how and if neural regions differ.

Gene expression arrays can provide some clues as to how different regions of the brain differ at the molecular level [4]. The difficulty arises in deciding what constitutes a brain region and how well the technology can detect any difference in global gene expression on a region-by-region basis. Most researchers use either Brodmann's definitions of cortical areas or gross anatomical features such as fissures or gyri to define regions.

Recently, Roth et al. [5] published a study comparing gene expression levels in a series of central nervous system (CNS) structures using microarray technology. The strengths of that paper were the quality of the sample used and the quality of the analysis performed. Their suggestions suffered somewhat in light of the fact that, for a marker to be region-specific or define a region, that marker had to be absent in any other brain region. Furthermore, the human cortex was considered like one region, while arguably, this region could be divided up into many different regions, suggesting that some of their potential markers may not be specific to the region they highlighted.

The purpose of this study was to further evaluate and confirm the claims of Roth et al. [5] in regions that overlap with our own work using our fully clinically characterized sample of normal controls. Using the same microarray platform in a different sample, we provide independent

support for a number of region-specific markers suggested by Roth et al. and provide additional evidence of the molecular specificity of brain regions in humans.

## Materials and methods

All subjects were of French–Canadian origin, a well-characterized genetically homogenous population [6], and died without a prolonged agonal state. All subjects were victims of motor vehicle accidents or cardiac arrest. For all 12 subjects (mean age=36 years; SEM=3.4 years), psychological autopsies were performed as described elsewhere [7] and, together with medical chart information, were free of any psychopathology or psychotropic medication.

The postmortem interval ranged from 20–36 h. Brains were extracted, sectioned, and snap frozen and stored at  $-80^{\circ}\text{C}$ . Cortical brain regions extracted were based on the Brodmann areas (BA): 4, 6, 8/9, 10, 11, 20, 21, 24, 29, 38, 44, 45, 46, and 47. Subcortical regions were also used, including the nucleus accumbens, hippocampus, and the amygdala. The final area examined was the cerebellum.

RNA from brain extractions was processed on the HG-U133AB and HG-U133 plus 2 chips. Analysis was done using MAS 5.0 and Genesis 2.0 (Gene Logic, Gaithersburg, MD). Subjects were excluded based on glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and B-actin 3':5' ratios. Only those subjects with B-actin 3':5' ratios above 0.3 and GAPDH ratios above 0.6 on every one of these markers were used. Cluster analysis was performed using average-linkage hierarchical cluster analysis with a correlation metric (Avadis; Strand Genomics, Redwood City, California).

To ensure the reliability of our data, we processed every sample on two separate microarray platforms. For every probe set, we generated a Pearson correlation coefficient between platforms across all subjects; only those probe sets with  $r^2$  values above 0.3 were used to confirm previously suggested markers.

To determine how specific these markers are across species, we compared positive markers from this confirmatory study to that of the mouse brain. We used the Allen brain atlas (<http://www.brain-map.org>) for marker comparison [8]. The Allen brain atlas is an online system in which global gene expression can be verified on a gene-by-gene basis throughout the entire mouse brain. Using only the confirmed genes, we entered the gene names and downloaded all images in as many planes as possible for every marker. We had specific criteria for anatomical regions when comparing our data to mouse data. The cerebellum is easily discerned due to its well-defined anatomy and specific structure. Both the nucleus accumbens core and shell were analyzed for this study. Using Paxinos and Watson criteria [9], the regions we inspected from the Allen brain atlas

correspond to: anteroposterior +1.3 to +1.6 mm; medio-lateral  $\pm 1.0$  to  $\pm 1.8$  mm from bregma; dorsoventral  $-6.8$  to  $-6.3$  mm from dura.

To assess expression level of neural markers in non-neural areas, we again relied on the Roth et al. [5] data, available at gene expression omnibus GSE3526. To compare these neural markers to other organs, we selected only those organs where there were four or more subjects, from five diverse areas and normalized using robust multiarray analysis [10].

## Results

Using the gene list from the study of Roth et al. [5] as our guide, we filtered microarray data for each of the genes in all 18 brain regions. Their study focused on the midbrain and lower CNS structures, whereas our study focused on cortical structures; however, both studies examined the cerebellum, the nucleus accumbens, and the hippocampus/amygdala complex. Therefore, we were able to verify their suggested markers in only these structures.

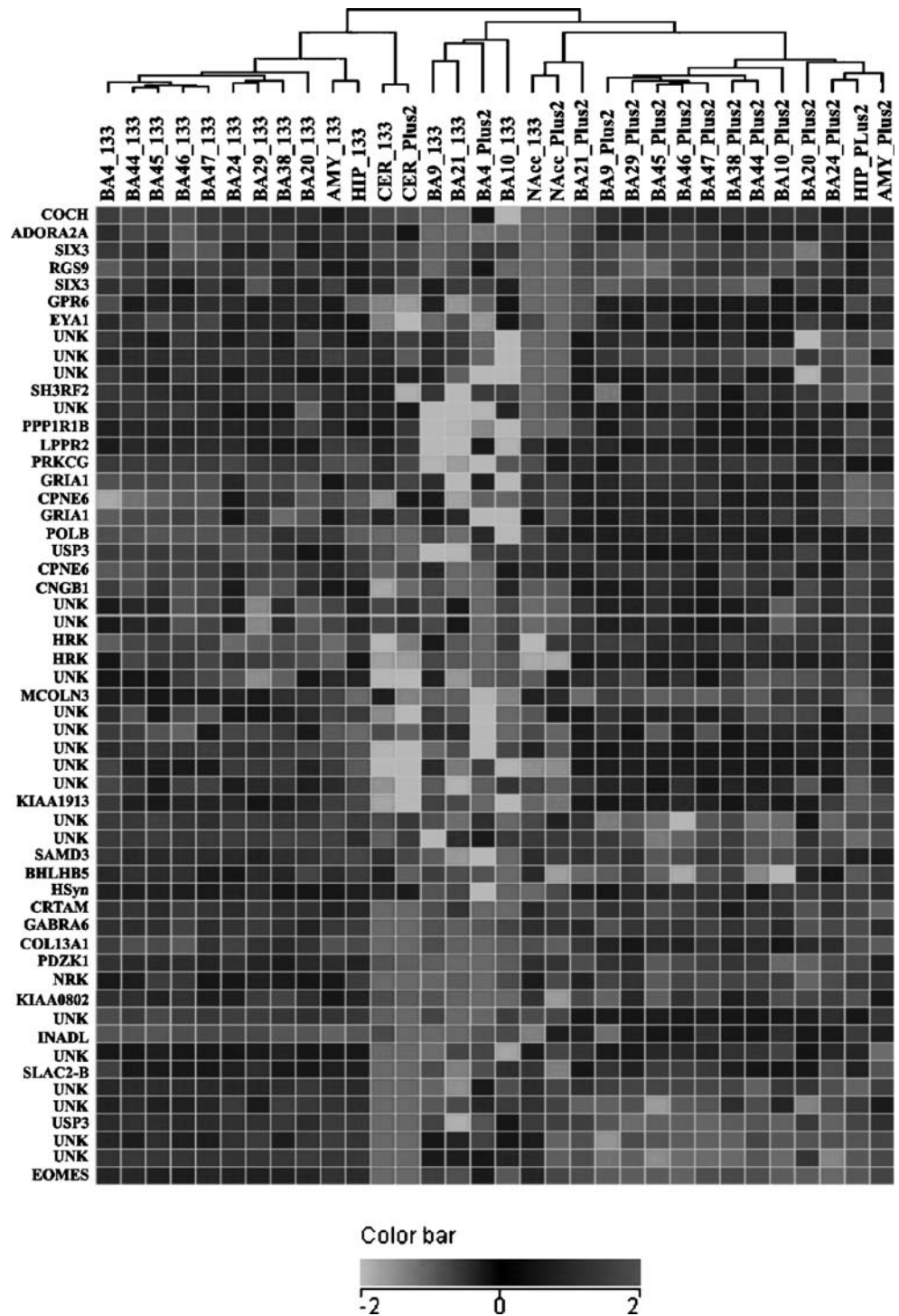
Using region specific markers proposed by Roth et al. [5], we examined the expression value for 54 potential markers. Across all subjects, we calculated a mean expression value for a given region for both the HG-U133 chip set and the HG-U133 plus 2 chip. Figure 1 is a heat map that demonstrates the results for this comparison. Following this first-pass analysis, we performed follow-up analysis on a gene-by-gene basis.

We had two criteria for whether a gene was considered a region-specific marker: First, the gene had to be expressed and called present (based on the  $p$  value generated through the comparison of a perfect match probe to a mismatch probe on the microarray, MAS 5.0) in only the area of interest and in no other region, or second, if the gene was present in more than the area of interest, that it was expressed at least threefold higher in the region of interest.

To begin our analysis on a gene-by-gene basis, we used a measure of quality control to ensure our own data was reproducible. To do this, we calculated the Pearson correlation coefficient between HG-U133AB and HG-U133 plus 2 chips. For this reason, only probe sets present on both chips were investigated. As shown in Table 1, results from both chips were highly reproducible as suggested by the high Pearson correlation coefficient between the chips.

Table 1 represents the findings from this study. Column 1 indicates the probe set ID and Column 2 is the gene symbol. Column 3 indicates whether or not this study replicates the findings from Roth et al. [5]. Those rows indicated by 'Yes' with a BA moniker indicate that the gene met criteria in all regions except for the area noted. The

**Fig. 1** Hierarchical clustering of potential region-specific markers from two different microarray platforms. Colors represent normalized values from 12 normal subjects. UNK refers to uncharacterized probe sets (unknown)



final column, PCC, represents the Pearson correlation coefficient between the HG-U133AB chip set and the HG-U133 plus 2 chip.

We note that some of the genes present in these tables are expressed in a small number of other neural regions; however, this graphical representation can be misleading. For example, in the case of GPR6 from the HG-U133 plus 2 chip in the NAcc, this probe set is 5.33 times greater than

the expression of the same probe set in HG-U133 plus 2 from BA21.

**Cross-organ marker expression**

We next asked how these brain expression markers compared to other organs in the human body. Using publicly

**Table 1** Probe sets (nucleus accumbens and cerebellum)

| Probe set ID             | Gene symbol | Replication? | PCC   |
|--------------------------|-------------|--------------|-------|
| <b>Nucleus accumbens</b> |             |              |       |
| 205229_s_at              | COCH        | Yes, BA 8/9  | 0.687 |
| 205013_s_at              | ADORA2A     | Yes, BA 8/9  | 0.342 |
| 242054_s_at              | SIX3        | Yes          | 0.995 |
| 214655_at                | GPR6        | Yes          | 0.979 |
| 230271_at                | Unknown     | Yes          | 0.660 |
| 239911_at                | Unknown     | Yes          | 0.813 |
| 243582_at                | SH3RF2      | Yes          | 0.737 |
| 227228_s_at              | (KIAA1509)  | Yes          | 0.962 |
| 229637_at                | Hsyn        | Yes          | 0.593 |
| 206518_s_at              | RGS9        | Yes, BA 10   | 0.440 |
| 233446_at                | Unknown     | Yes          | 0.826 |
| 225165_at                | PPP1R1B     | Yes          | 0.902 |
| <b>Cerebellum</b>        |             |              |       |
| 206914_at                | CRTAM       | Yes, BA 10   | 0.759 |
| 207182_at                | GABRA6      | Yes          | 0.974 |
| 236793_at                | Unknown     | Yes          | 0.987 |
| 203616_at                | POLB        | Yes, BA 8/9  | 0.873 |
| 235221_at                | Unknown     | Yes          | 0.998 |
| 211343_s_at              | COL13A1     | Yes, BA 10   | 0.611 |
| 233435_at                | Unknown     | No           | 0.606 |
| 214734_at                | SLAC2-B     | Yes, BA 8/9  | 0.934 |
| 243509_at                | Unknown     | Yes          | 0.949 |
| 221654_s_at              | USP3        | No           | 0.900 |
| 227971_at                | NRK         | Yes          | 0.845 |
| 231776_at                | EOMES       | Yes          | 0.987 |
| 239344_at                | Unknown     | Yes          | 0.985 |
| 231051_at                | Unknown     | Yes          | 0.986 |

PCC is Pearson correlation coefficient

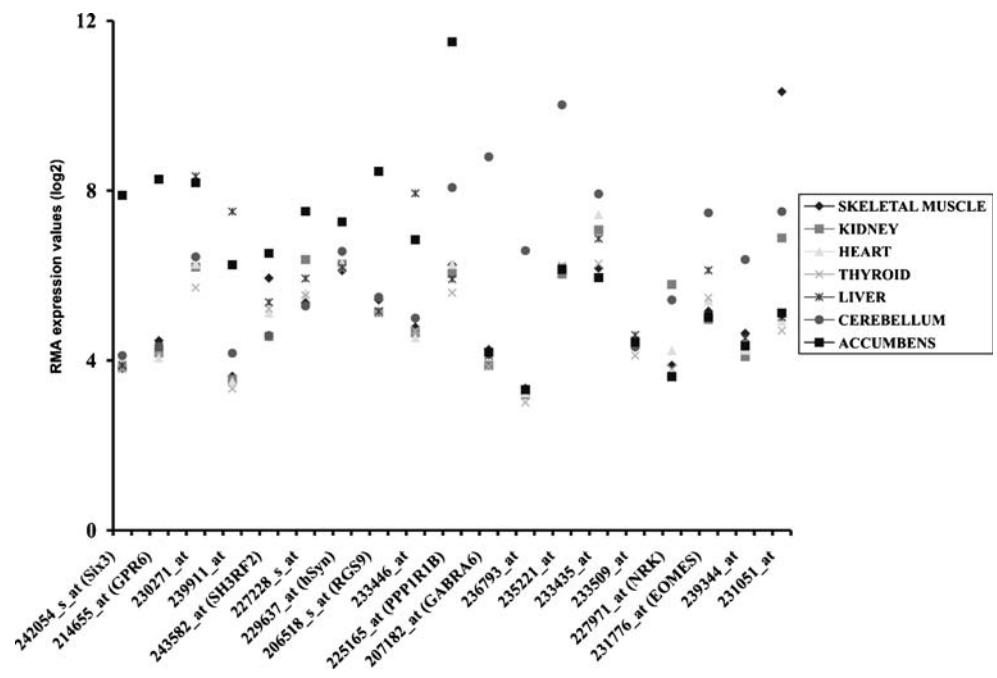
available data (Gene expression omnibus; GSE3526; the Roth dataset), we analyzed the expression of all markers in skeletal muscle, liver, kidney (cortex), heart (atrium), and thyroid gland (Fig. 2). For comparison reasons, we have also included the expression values from the cerebellum and nucleus accumbens from the Roth dataset. The raw data for this figure can be found in Supplemental Table 1.

### Cross-species marker expression

We next asked how the expression of the markers that were identified in both the Roth study and our own study compared to the expression of the same markers in the mouse brain. To assess this, we used information available publicly from Allen Brain Institute. We downloaded information from only those markers that had known names and validated in both our own study and the Roth study. In the nucleus accumbens, we searched for Six3, GPR6, SH3RF2, hSyn, and PPP1R1B and from the cerebellum: gamma-aminobutyric-acid A receptor alpha-6 (GABRA6), Nik-related kinase (NRK), and eomesodermin (EOMES).

In mouse brain, GPR6 showed a remarkably similar pattern to that of humans: Expression was restricted to the NAcc core and shell, although the signal seemed to encompass slightly more than just the NAcc. No other signal for this gene could be detected anywhere else in the mouse brain confirming its high specificity in the NAcc. According to the Allen brain atlas, both SH3RF2 and PPP1R1B are ubiquitously expressed in mouse brain, including cerebellum, hippocampus, and frontal cortex. Both GABRA6 and NRK showed excellent specificity to

**Fig. 2** Expression of brain-specific markers in five non-neural organs as well as the cerebellum and accumbens



the cerebellum, although both were faintly detected in the hippocampus. Finally, EOMES seemed to be ubiquitously expressed in mouse brain, although the irregular staining pattern suggests that this may have been a quality problem.

## Discussion

This study investigated whether previously suggested molecular markers of brain regions are truly region specific using an independent genetically homogenous population [6], a different extraction method, a different analysis method, and two different chip types from Affymetrix for all the subjects. Having a genetically homogenous population has the advantage of eliminating gene expression differences due to ethnic background and shown to have important effects in global gene expression measurements even when looking at cells from individuals from different population groups [11]. In general, we found that, in both the cerebellum and the nucleus accumbens, the markers previously suggested are indeed accurate and that the suggested markers' expression levels are highly reproducible between the two chipsets used in this study.

The studies were different in some ways. Roth et al. [5] grouped the hippocampus with the amygdala for their analysis, which may explain the lack of replication in this area. As well, Roth et al. grouped the nucleus accumbens with the putamen, suggesting that the markers replicated here are likely specific to both the nucleus accumbens and the putamen. Both our study and that of Roth et al. used the HG-U133 plus 2 chip, whereas we made an effort to reproduce our findings using a different platform (HG-U133AB).

In the nucleus accumbens, some of the functions of these genes have been characterized hSyn (RIC8A) is thought to regulate neuronal transmitter release and interacts with Galpha [12]. PPP1R1B (DARPP-32) is involved in cyclic adenosine monophosphate and dopamine signaling and has a known role in dopaminergic signaling in the striatum [13]. This clearly is not a marker specific to the NAcc. As neither our study nor the Roth study verified the caudate nucleus or the putamen individually, it was foreseeable that this was detected as a specific marker of the NAcc. This marker was detected throughout the mouse brain in the Allen brain atlas. Six3 has a known developmental role in patterning of the anterior neuroectoderm [14], although its role in adult brain is not known. We could find no studies that explored functions of either SH3RF2 or GPR6, an orphan g-protein coupled receptor.

GABRA6, specifically expressed only in the cerebellum, has already been shown to be preferentially expressed in the postnatal cerebellum [15]. Interestingly, a polymorphism (Pro385Ser) of this GABA A receptor subunit has also been

associated with depression related traits [16], and given its cerebellum-specific expression, this is a point of some curiosity and suggest a possible implication of the cerebellum in depression. EOMES, a DNA-binding protein that may play a role in immunity [17], and NRK have a role in the cytoskeleton, particularly actin dynamics [18].

The combined power of both the Roth et al. [5] study and this one strongly suggest that these genes are truly markers of the brain regions described. Together, these studies have essentially examined almost the entire cortical region and lower CNS structures. Of note is the lack of gene expression analysis from the occipital and parietal lobes that could prove to later express some of these markers. Finally, much debate has occurred over the reliability of microarray data [19]. This study, using microarray data and performed completely independently from the Roth et al. group, suggests that microarray data can be a reliable measure of gene expression when properly processed and analyzed.

**Acknowledgment** Carl Ernst and Adolfo Sequeira contributed equally to this work.

## References

- Rakic P (1988) Specification of cerebral cortical areas. *Science* 241(4862):170–176
- Penfield W (1957) Some observations on the functional organization of the human brain. *Ideggyogyasz Szle* 10(5–6):138–141
- Gravel C, Leclerc N, Rafrafi J, Sasseville R, Thivierge L, Hawkes R (1987) Monoclonal antibodies reveal the global organization of the cerebellar cortex. *J Neurosci Methods* 21(2–4):145–157
- Murray KD, Choudary PV, Jones EG (2007) Nucleus- and cell-specific gene expression in monkey thalamus. *Proc Natl Acad Sci U S A* 104(6):1989–1994
- Roth RB, Hevezi P, Lee J, Willhite D, Lechner SM, Foster AC, Zlotnik A (2006) Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. *Neurogenetics* 7(2):67–80
- Simard LR, Prescott G, Rochette C, Morgan K, Lemieux B, Mathieu J, Melancon SB, Vanasse M (1994) Linkage disequilibrium analysis of childhood-onset spinal muscular atrophy (SMA) in the French–Canadian population. *Hum Mol Genet* 3(3):459–463
- Dumais A, Lesage AD, Alda M, Rouleau G, Dumont M, Chawky N, Roy M, Mann JJ, Benkelfat C, Turecki G (2005) Risk factors for suicide completion in major depression: a case-control study of impulsive and aggressive behaviors in men. *Am J Psychiatry* 162(11):2116–2124
- Lein ES, Hawrylycz MJ, Ao N et al (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445(7124):168–176
- Paxinos G, Watson C (1998) *The rat brain in stereotaxic coordinates*. Academic, San Diego
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4(2):249–264
- Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG (2007) Common genetic variants account for differ-

- ences in gene expression among ethnic groups. *Nat Genet* 39(2): 226–231
12. Klattenhoff C, Montecino M, Soto X, Guzman L, Romo X, Garcia MA, Mellstrom B, Naranjo JR, Hinrichs MV, Olate J (2003) Human brain synembryon interacts with Gsalpha and Gqalpha and is translocated to the plasma membrane in response to isoproterenol and carbachol. *J Cell Physiol* 195(2):151–157
  13. Ouimet CC, Greengard P (1990) Distribution of DARP P-32 in the basal ganglia: an electron microscopic study. *J Neurocytol* 19(1): 39–52
  14. Wallis DE, Roessler E, Hehr U, Nanni L, Wiltshire T, Richieri-Costa A, Gillessen-Kaesbach G, Zackai EH, Rommens J, Muenke M (1999) Mutations in the homeodomain of the human SIX3 gene cause holoprosencephaly. *Nat Genet* 22(2):196–198
  15. Wisden W, Korpi ER, Bahn S (1996) The cerebellum: a model system for studying GABAA receptor diversity. *Neuropharmacology* 35(9–10):1139–1160
  16. Sen S, Villafuerte S, Nesse R, Stoltenberg SF, Hopcian J, Gleiberman L, Weder A, Burmeister M (2004) Serotonin transporter and GABAA alpha 6 receptor variants are associated with neuroticism. *Biol Psychiatry* 55(3):244–249
  17. Intlekofer AM, Takemoto N, Wherry EJ, Longworth SA, Northrup JT, Palanivel VR, Mullen AC, Gasink CR, Kaech SM, Miller JD, Gapin L, Ryan K, Russ AP, Lindsten T, Orange JS, Goldrath AW, Ahmed R, Reiner SL (2005) Effector and memory CD8+ T cell fate coupled by T-bet and eomesodermin. *Nat Immunol* 6(12):1236–1244
  18. Nakano K, Kanai-Azuma M, Kanai Y, Moriyama K, Yazaki K, Hayashi Y, Kitamura N (2003) Cofilin phosphorylation and actin polymerization by NRK/NESK, a member of the germinal center kinase family. *Exp Cell Res* 287(2):219–227
  19. Draghici S, Khatri P, Eklund AC, Szallasi Z (2006) Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet* 22(2):101–109