International Annals of Science
ISSN: 2456-7132
Volume 9, Issue 1, pp. 86-99, 2020
DOI: https://doi.org/10.21467/ias.9.1.86-99

RESEARCH ARTICLE

Differential Expression Analysis on Schizophrenia Dataset Suggests Pseudogene RNU6-505P as under Selective Pressure

Emilio Mastriani[1,2*], Alexey V. Rakov[4,5], Shu-Lin Liu[1,2,3]

[1] Systemomics Center, College of Pharmacy, Genomics Research Center, State-Province Key Laboratories of Biomedicine-Pharmaceutics of China, Harbin Medical University, Harbin, China.
[2] HMU-UCCSM Centre for Infection and Genomics, Harbin Medical University, Harbin, China.
[3] Department of Microbiology, Immunology and Infectious Diseases, University of Calgary, Calgary, Canada.
[4] Laboratory of Molecular Epidemiology and Ecology of Pathogenic Bacteria, Somov Institute of Epidemiology and Microbiology, Vladivostok, Russia.
[5] Department of Medical Biology and Biotechnology, School of Biomedicine, Far Eastern Federal University, Vladivostok, Russia.

* Corresponding author email: emiliomastriani@hrbmu.edu.cn

Received: 22 October 2019 / Revised: 31 December 2019 / Accepted: 10 January 2020 / Published: 25 January 2020

ABSTRACT

Schizophrenia is one of the 15 leading causes of disability worldwide. About 1% of the global population has schizophrenia, with 10% of premature mortality chance. Schizophrenia is therefore associated with significant health, social and economic concerns. In this context, thalamus and striatum areas play important roles as much in schizophrenia as processing information before reaching the conscious thought: step happening soon in the creativity action. Creativity is defined by psychological scientists as the generation of ideas or products that are both original and valuable. Creativity relies on imagination and this fundamental human ability remains understudied in comparison to other important psychological phenomena. It is natural to ask whether the gene expression profiling of samples from schizophrenic patients could highlight the activity of some genes specific to humans. Microarray analysis of the dataset GSE25673 revealed that the pseudogene RNU6-505P is expressed differentially in schizophrenic samples and correlates to CYP26A1, ARHGAP18, TSPAN12, HEY2 and TMEM132A genes. Ontological analysis showed that the RNU6-505P pseudogene is involved in brain development and certain neurological pathologies. Evolutionary analysis showed that the AGA 3-nucleotide sequence of RNU6-505P has been under positive selective pressure. Finally, the 1-nucleotide mutation prediction test revealed that variations on the AGA nucleotides could be fatal to the RNA structure of the sequence. We conclude that differential expression of the RNU6-505P pseudogene can be valid to diagnose schizophrenia and the RNU6-505P pseudogene may have a relevant function in the cerebral development and in the divergent evolution of humans from apes.

Keyword: Differential expression, Brain, Evolution, Microarray, Schizophrenia, Pseudo-gene, RNU6-505P, Secondary structure prediction

1 Introduction

Finding a starting point that suggests how the human being has differentiated from common ancestors with apes is certainly a very fascinating and complex topic. The innate abilities to create and imagine are the prerogatives of the human species. Creativity (to create art, invent tools, think scientifically, etc.) is a remarkable activity normally operated by humans. And this,

at least as far as we know at the moment, could distinguish the human species from its ancestors and close relatives. The investigation into the origins of human creativity may provide important information about the evolutionary process. Previous works [1-3] have tried to find a relationship between the state of *"madness"* and *"creativity"*. On the other hand, regions of the brain (e.g., the striatum and the thalamus) are principal actors both in information processing before it reaches the state of consciousness (soon step of creative activity) and in the case of patients with schizophrenia. Through the investigations of the anatomy of the brains of primates [4], the nature of the interconnected tissues of the cerebellum to the multiple motor cortices suggested that the thalamus fulfills a key function in providing the specific channels from the basal ganglia and cerebellum to the cortical motor areas.

Differential gene expression in schizophrenia brain cells may elucidate which genes might be involved in disease progression, such as those that are expressed statistically differently in schizophrenia patients [5]. One GWAS study showed that 185 genes may have associations with schizophrenia [6]. The involved genes may be up- or down-regulated depending on a stage of disease [7]. Some recent studies on increased pseudogene expression in schizophrenia-derived cell lines showed that the NDUFV2P1 pseudogene contributes to mitochondrial complex I deficits [8].

In this study, we hypothesized that genetic analysis of schizophrenia may lead to clues to the evolutionary process of human and estimated possible selective pressures on specific nucleotides of samples from schizophrenic patients, using the RNU6-505P pseudogene as a marker for schizophrenia, with a focus on its postulated activity in the development of the human brain and involvement in different brain functions/diseases. The selective pressures on a specific region preceding the RNU6-505P pseudogene containing the AGA sequence seems to be related to its secondary structure, suggesting that RNU6-505P may have played important roles in the human evolutionary process.

2 Materials and Methods

2.1 Differential Cluster Analysis

We started the research with the microarray analysis of the dataset GSE25673 [9], comparing control and schizophrenic hiPSC-derived neurons. In order to be sure about the quality of the information, we conducted all the preliminary steps using the Oligo [10] package into Bioconductor 3.4 [11]. As could be seen in Figure 1, the arrays are nicely centered around Relative Log Expression (RLE), with approximately equal box sizes and no quality control problems. From the Normalized Unscaled Standard Error (NUSE), it appears that the arrays are reasonably centered around the median (NUSE = 1) and but do not appear to present any quality control {Emilio: please change it back if my "correction" was not correct}. Applying the RMA (robust multichip average) background adjustment, normalization and summarization, we observe (Figure 2) that the boxplots of all arrays are nicely aligned (A) and that the low regression line (B) provides an evidence that the applied procedures were adequate.

Then we used Limma [12] package to calculate a moderated t-statistic with shrunken standard deviation for each gene. An empirical Bayes method was used to shrink the variance of each gene towards a common value for all the genes. Note that this has done to lower the influence of very low or very high standard deviations on the t-test. The subsequent cluster analysis, has been conducted using the mclust [13] package.

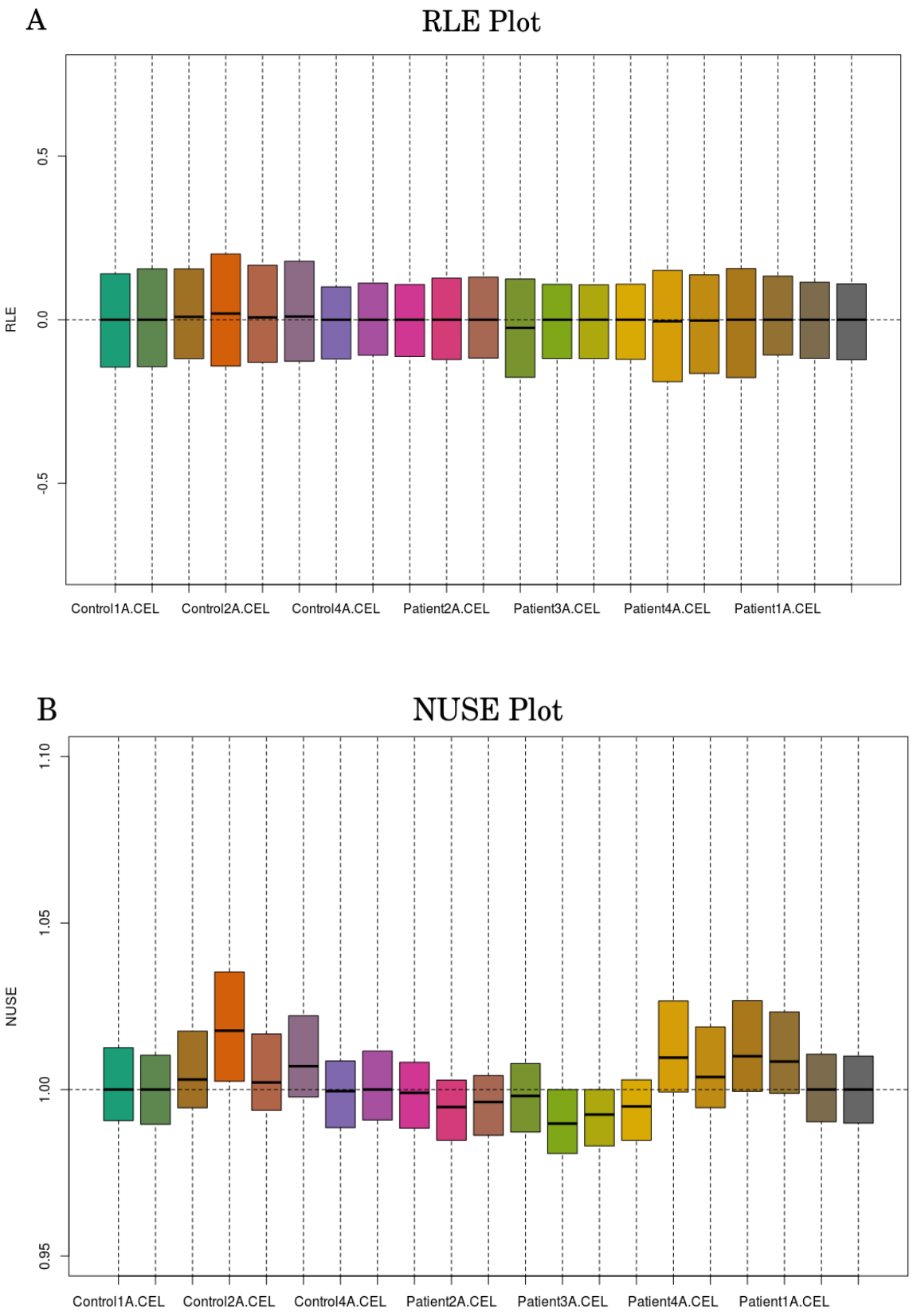

Figure 1: *RLE and NUSE boxplots after the RMA*

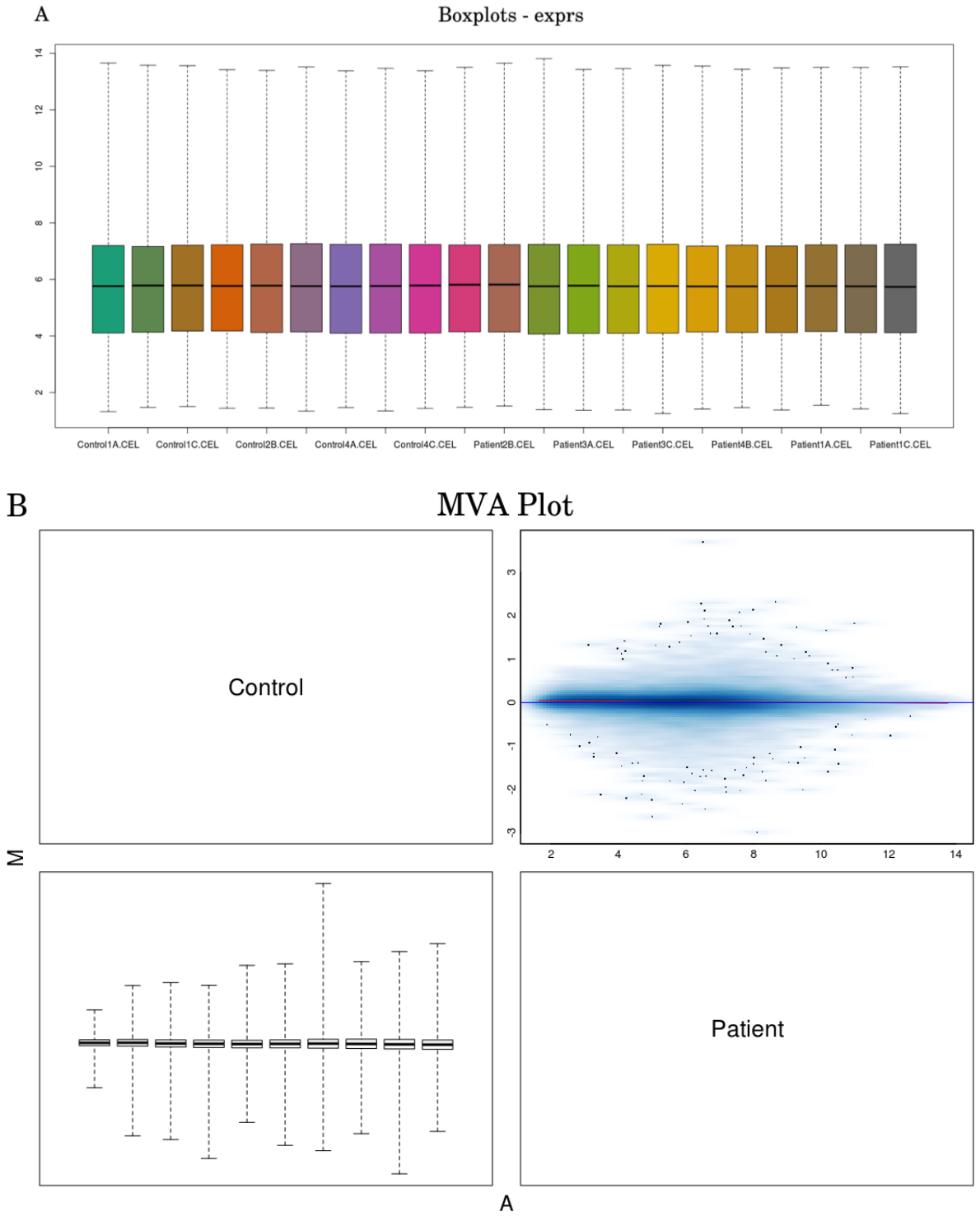

Figure 2: *Boxplots and MA plot after the RMA*

2.2 Correlation Analysis

Limma package was also used to calculate the Pearson correlation. We utilized RPISeq tool [14] to estimate the goodness of the interaction between the RNU6-505P pseudogene and its correlated genes. This method, based on machine-learning approach, can readily predict RNA-protein interactions using only sequence information. RPISeq predictions are based on Random Forest (RF) or Support Vector Machine (SVM) classifiers trained and tested on 2 non-redundant benchmark datasets of RNA-protein interactions, RPI2241 and RPI369, extracted from PRIDB [15]. Also, we used the NCBI BLAST tool to make comparative analysis [16].

2.3 Gene Network and Ontological Analysis

In order to obtain an exhaustive gene network concept, we preliminarily used GeneAnswers [17] package for investigating the functions assumed by the genes correlated to RNU6-505P. Since we worked with the EGAN tool [18] on Entrez Human Dataset [19] to collect information from all available genes-functions-ontology databases. We addressed OMIM [20] to investigate the pathologies where "correlated" genes are involved. ATLAS database [21] has accessed to get the expression level of genes in prenatal human brain.

2.4 Evolutionary Analysis

Ete3 [22] toolkit (based on CodeML program from PAML package [23]) supported the research of selective pressure defined by the ratio dN/dS (synonymous rate/non-synonymous rate). We considered the 345bp sequence (119bp before and after the RNU6-505P pseudogene) of human, chimpanzee, gorilla and mouse. We applied the bsA1 as null model and bsA as alternative model (different dN/dS) for testing the positive pressure on sites on specific branches [24-26].

2.5 Secondary Structure Mutation Prediction

Prediction of the secondary structure of RNU6-505P *"extended"* sequence (345bp) has obtained by RNAFold [27] program. The same tool was used to calculate the secondary structure of the "extended" RNU6-505P pseudogene in case of single nucleotide mutation.

3 Results and Discussion

3.1 Differential Expression of RNU6-505P and Cluster Analysis

Differential analysis of the dataset revealed that RNU6-505P belongs to the subset of genes whose $logFC > 2$, as we can see by the volcano plot in Figure 3 RNU6-505P looks up regulated same as other genes in the red cluster. In fact, taking in consideration the top 100 genes sorted by the adjusted

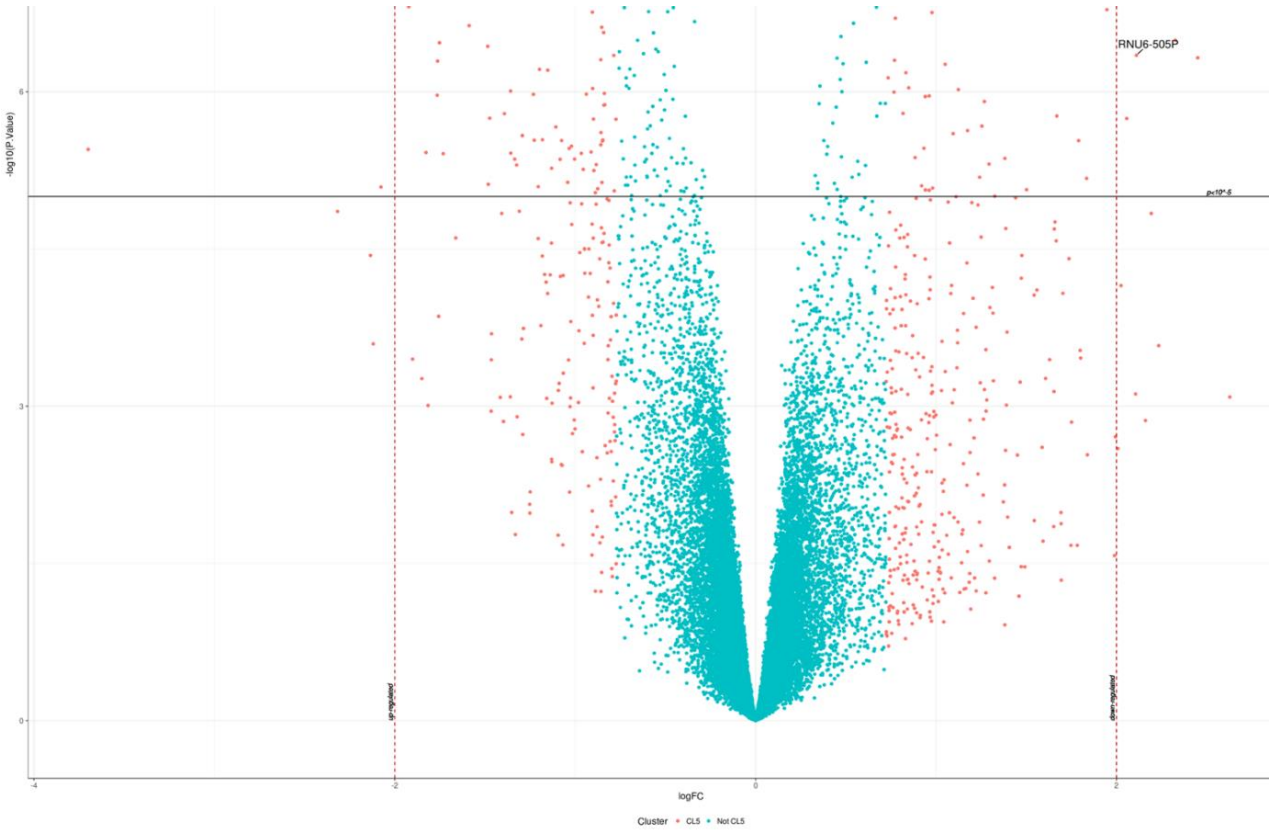

Figure 3. *Volcano plot of the dataset. In red colour are genes in CL5; dashed red lines indicate up and down regulated genes; also, the black line with $p < 10^{-5}$ is shown.*

P value, cluster analysis shown (Figure 4) that 63% of them belong to a small (568 of 33297) group (CL5). As we can see by the coming short Table 1, the highlighted genes have adjusted P value $\leq 1.2E^{-0.5}$ and all of them show a higher level of differential expression on schizophrenic patients, such that they are placed in the *"erupting"* portion of the volcano diagram (Figure 3).

Table 1: *Adjusted P values of the top 20 genes*

SYMBOL	Adj.P.Value	Cluster
CHST9	2.5E-0.7	CL5
PDE4D	2.E-0.7	CL5
C4B	1.2E-0.6	CL5
NET1	1.5E-0.6	CL5
GLI3	1.5E-0.6	CL5
ZIC1	1.6E-0.6	CL5
SLC34A2	1.6E-0.6	CL5
CYP26A1	1.8E-0.6	CL5
TSPAIN2	1.8E-0.6	CL5
PRCP	1.9E-0.6	Not CL5
ALDH1L1	2.0E-0.6	CL5
S100A16	3.5E-0.6	CL5
UCP2	6.8E-0.6	Not CL5
PRDM8	7.8E-0.6	Not CL5
WASF2	7.8E-0.6	Not CL5
MATN2	7.8E-0.6	CL5
HEY2	7.8E-0.6	CL5
KCND3	8.1E-0.6	CL5
ARAP2	1.2E-0.5	CL5

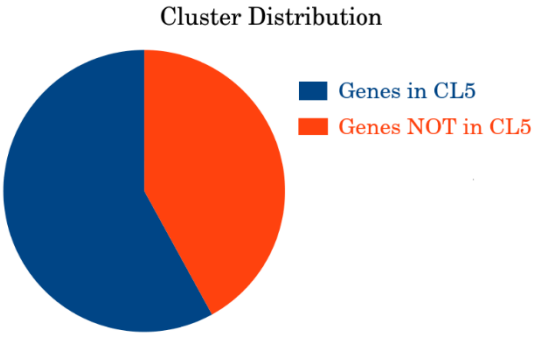

Figure 4: *Distribution of the top 100 genes over 5 clusters*

3.2 Correlation of RNU6-505P to Specific Genes

In attempt to conceive a valid hypothesis about the probable rule carried out by the RNU6-505P gene, we calculated the Pearson correlation and its relative significance for all genes in CL5. The Table 2 shows the list of the top genes with $Pearson\ coefficient\ of\ correlation \geq |0.76|$ (positive and negative) and $Pearson\ significance \leq |5.97E^{-0.05}|$. Reporting the best 5 genes (up/down regulated) on the volcano plot (Figure 5), we can observe their level of differential expression. In detail, we detected the following genes differentially expressed and strongly correlated to RNU6-505P:

- CYP26A1and ARHGAP18 are up-regulated (same as RNU6-505P)
- TSPAN12, HEY2 and TMEM132A are down-regulated (opposite to RNU6-505P)

Table 2: *List of genes positively/negatively correlated to RNU6-505P*

Down Regulated			Up Regulated		
SYMBOL	**Pearson Corr.**	**Pearson Sign.**	**SYMBOL**	**Pearson Corr.**	**Pearson Sign.**
HEY2	-0.9368	4.1E-1.0	ZIC4	0.94881	5.9E-11
TIMP3	-0.9210	3.2E-0.9	CHMP1B	0.94322	1.5E-10
S100A16	-0.9191	4.0E-0.9	CYP26A1	0.9464	1.7E-10
ALDOC	-0.9184	4.3E-0.9	RNU6-1169P	0.93727	3.9E-10
CYYR1	-0.9101	1.1E-0.8	ARHGAP18	0.91637	5.5E-0.9
TSPAN12	-0.8949	4.4E-0.8	TNFRSF11B	0.91431	6.8E-0.9
KAL1	-0.8874	8.2E-0.8	CHRM3	0.90925	1.2E-0.8
TMEM132A	-0.8798	1.5E-0.7	WDR91	0.90431	1.9E-0.8
ASTN1	-0.8756	2.0E-0.7	FIGN	0.90248	2.2E-0.8
C4B	-0.7944	1.66E-0.05	ZIC1	0.88082	1.37E-0.07
			CD36	0.85757	6.77E-0.07
			MATN2	0.78621	2.38E-0.05
			FAM46A	0.78182	2.84E-0.05
			RIMS2	0.76195	5.97E-0.05

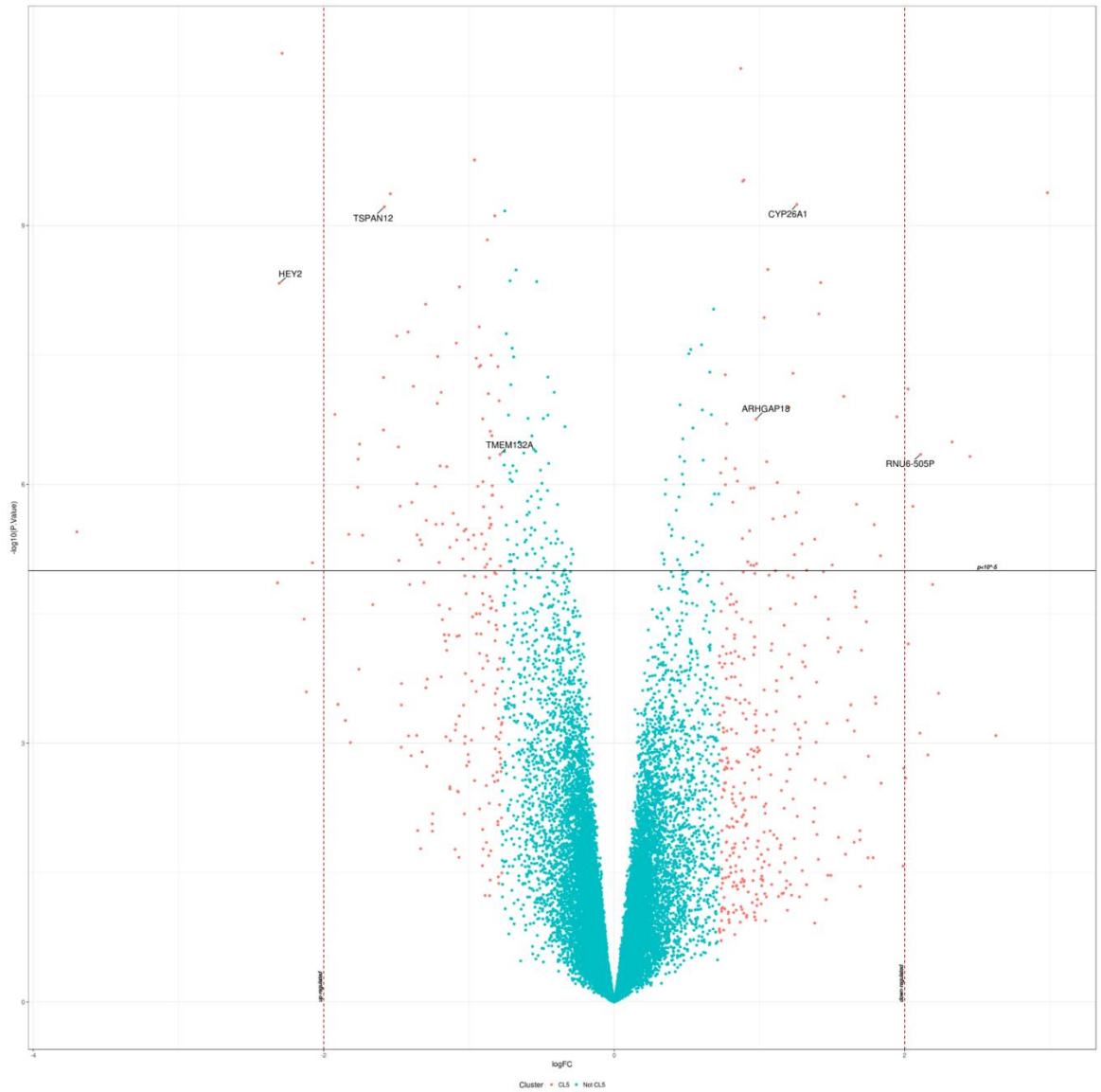

Figure 5: *Top 5 genes strongly correlated to RNU6-505P and differentially expressed*

The Table 3 reports the RF and SVM classifiers for the interaction between the RNA sequence of RNU6-505P and the proteins encoded by the genes in Table 2. Outcomes show that genes directly correlated to RNU6-505P have $SVM < 0.8$, while the range $0.6 < SVM < 0.8$ is identified for those "inversely" correlated, suggesting a partitioning between the 2 sets.

Is the prediction of RNA-protein interactions in the case of human the same as other species? The Figure 6 shows the SVM classifier in case of human, gorilla, chimpanzee and mouse. In details, in the case of human we took in consideration the SVM values calculated for the interaction between RNA6-505P and all mentioned proteins (Table 3); considering the case of the other species: we calculated the SVM value between the RNU6-505P ortholog and the ortholog of proteins considered previously (Table 3). Results advise that the RNA6-505P/Proteins interacting could be human specific, because in most of case the SVM value is higher for human. And also, this analysis hints that could be possible to consider the *"mouse model"* to test the real connections between RNU6-505P and proteins whose SVM values are greater than 0.860 and coincident to the human case (C4B, TIMP3, ZIC4, WDR91). It's interesting to note that all of them are well expressed in the *choroid plexus*, while the human specific (CYP26A1 and ARHGAP18) are respectively strongly expressed in *cerebellum* and *ganglion*.

Table 3: *RNA-protein interaction prediction based on sequences. In green the upregulated genes, while in grey the downregulated*

UNIPROT ID	RF	SVM	Gene	DESC
Q9UQ26|RIMS2_HUMAN	0.75	0.961	RIMS2	Regulating synaptic membrane exocytosis protein 2
P0C0L5|CO4B_HUMAN	0.65	0.931	C4B	Complement C4-B
O43174-2|CP26A_HUMAN	0.65	0.924	CYP26A1	Cytochrome P450 26A1
Q8N9L1|ZIC4_HUMAN	0.6	0.922	ZIC4	Zinc finger protein ZIC 4
P20309|ACM3_HUMAN	0.65	0.919	CHRM3	Muscarinic acetylcholine receptor M3
A4D1P6|WDR91_HUMAN	0.65	0.91	WDR91	WD repeat-containing protein 91
O14525|ASTN1_HUMAN	0.5	0.91	ASTN1	Astrotactin-1
Q8N392|RHG18_HUMAN	0.6	0.893	ARHGAP18	Rho GTPase-activating protein 18
P23352|KALM_HUMAN	0.65	0.87	KAL1	Anosmin-1
Q5HY92|FIGN_HUMAN	0.55	0.863	FIGN	Fidgetin
Q24JP5|T132A_HUMAN	0.75	0.85	TMEM132A	Transmembrane protein 132A
P16671|CD36_HUMAN	0.6	0.839	CD36	Platelet glycoprotein 4
P35625|TIMP3_HUMAN	0.6	0.78	TIMP3	Metalloproteinase inhibitor 3
O95859|TSN12_HUMAN	0.55	0.78	TSPAN12	Tetraspanin-12
O00339|MATN2_HUMAN	0.8	0.769	MATN2	Matrilin-2
O00300|TR11B_HUMAN	0.65	0.766	TNFRSF11B	Tumor necrosis factor receptor superfamily member 11B
Q96IP4|FA46A_HUMAN	0.65	0.736	FAM46A	Protein FAM46A
P09972|ALDOC_HUMAN	65	0.73	ALDOC	Fructose-bisphosphate aldolase C
Q7LBR1|CHM1B_HUMAN	0.5	0.726	CHMP1B	Charged multivesicular body protein 1b
Q15915|ZIC1_HUMAN	0.65	0.72	ZIC1	Zinc finger protein ZIC 1
Q9UBP5|HEY2_HUMAN	0.65	0.676	HEY2	Hairy/enhancer-of-split related with YRPW motif protein 2
Q96FQ6|S10AG_HUMAN	0.45	0.65	S100A16	Protein S100-A16
Q96J86|CYYR1_HUMAN	0.6	0.56	CYYR1	Cysteine and tyrosine-rich protein 1

3.3 Expression of RNU6-505P into Neuro Pathologies

The diagram reported in Figure 7 points out the most important properties about the biological function and component for CYP26A1, ARHGAP18, TSPAIN12, HEY2 and TMEM32 genes. In this representation, the link between two genes (direct or mediated) shows the existing biological correlation and the arrows from RNU6-505P stress the possible links to the correlated genes. This schema highpoints the reasonable engagements that RNU6-505P has in many biological functions.

Another way to figure out the possible functions of RNU6-505P is to investigate the identified pathologies of its correlated genes. The OMIM diagram in Figure 8 depicts the disease network of the genes strongly correlated to RNU6-505P. As we suspected, most of the diseases are generally neurological, while ARHGAP18 is directly linked to schizophrenia, confirming again our hypothesis of correlation.

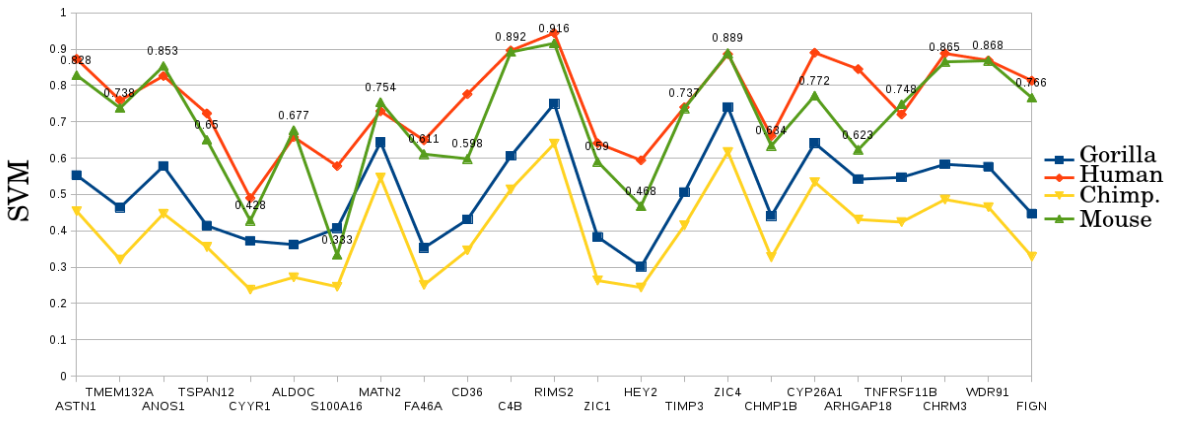

Figure 6: *Comparison of RNA-protein interaction prediction*

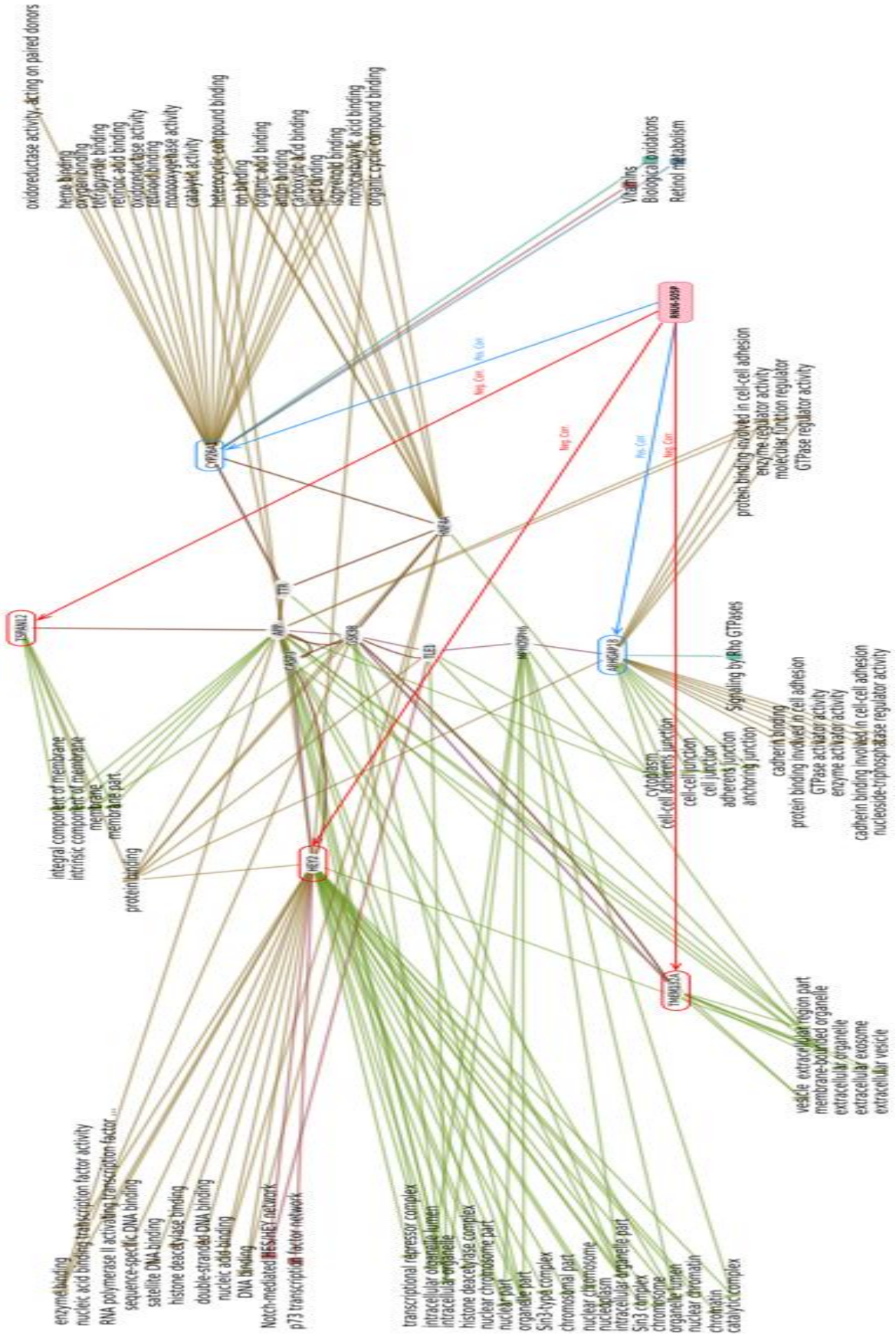

Figure 7: *Gene Network Concept of RNU6-505P and its strongly correlated genes that are differentially expressed. Blue colour means Positive Correlation, red means Negative Correlation*

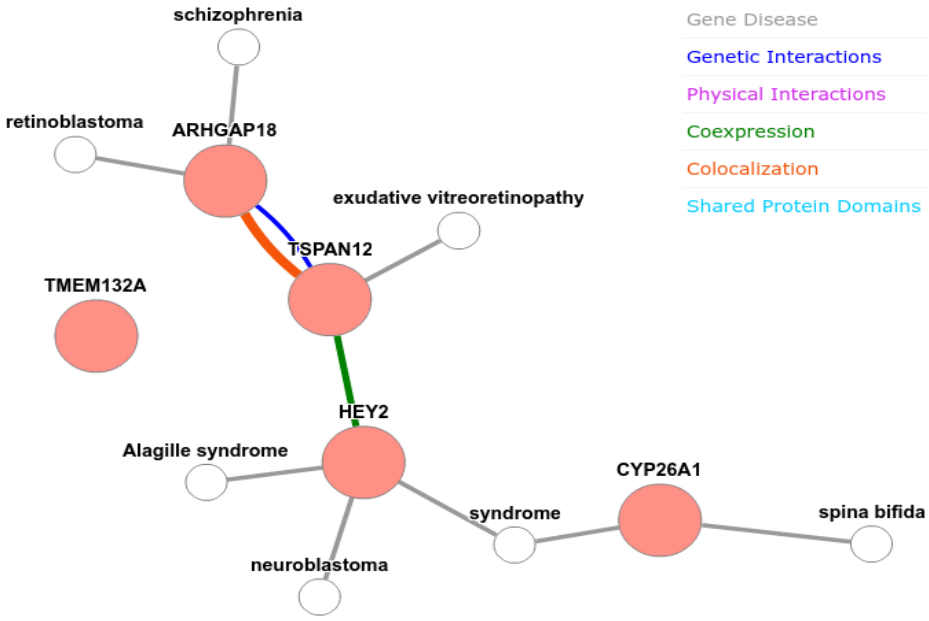

Figure 8: *Disease gene network*

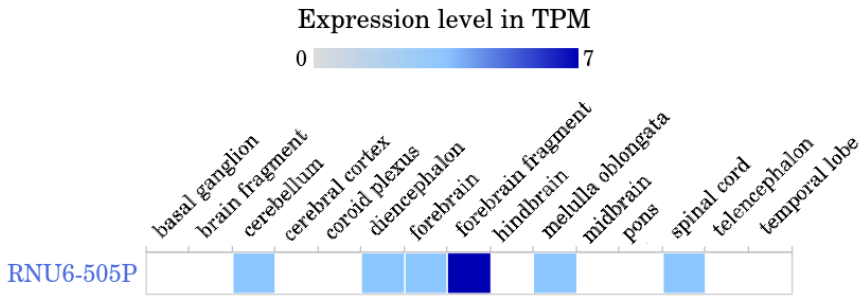

Figure 9: *mRNA expression in normal human tissues for RNU6-505P gene during the prenatal development*

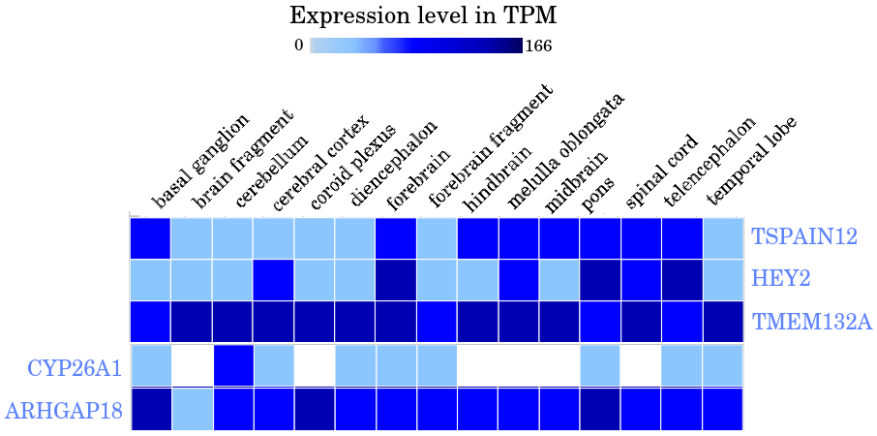

Figure 10: *mRNA expression of correlated genes*

Figure 9 reports the level of mRNA expression in normal tissue for RNU6-505P (Genecards website [28]) during the prenatal human brain development, while Figure 10 depict the same information for its correlated genes. Considering that the expression level of mRNA RNU6-505P is stronger on *brain* and *cerebellum* tissues (Figure 9) and that we got the same evidence for the genes strongly correlated to RNU6-505P from prenatal human brain development (Human Developmental Biology Resource, ENSG00000255112 from Atlas) at the 9th week after conception, results suggest that RNU6-505P could really play a crucial rule in the brain development.

3.4 Positive Selective Pressure of Nucleotides on RNU6-505P (AGA site)

Firstly, we focused on the human evolution using his branch like foreground and all others like background. Secondly, we have performed the same on the chimpanzee case. By the comparison of the two evolutionary models, we observe (Figure 11) that some regions seem to be under positive pressure (peaks), but only someone are human specific: TAT, AGA, GCC and ACA (red arrows). This result points out that, for some reason, the RNU6-505P pseudogene has been object of evolution pressure for human.

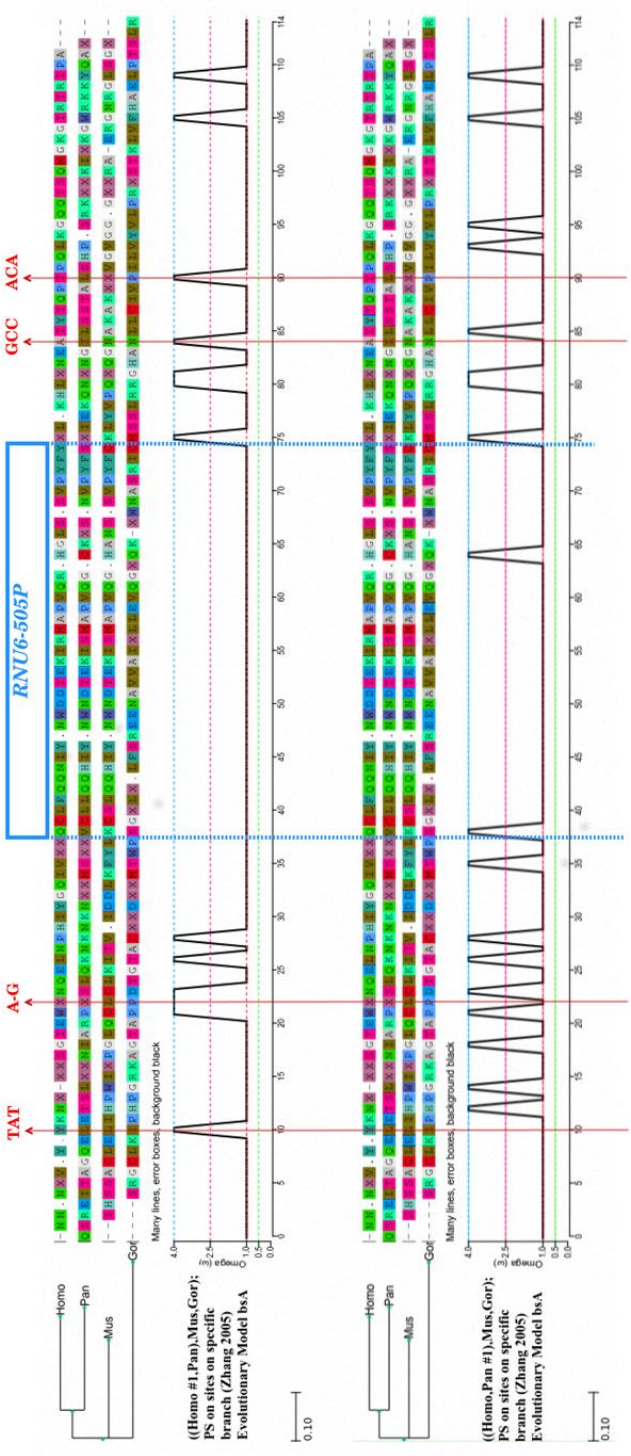

Figure 11: *Selective pressure of the extended region near to RNU6-505P*

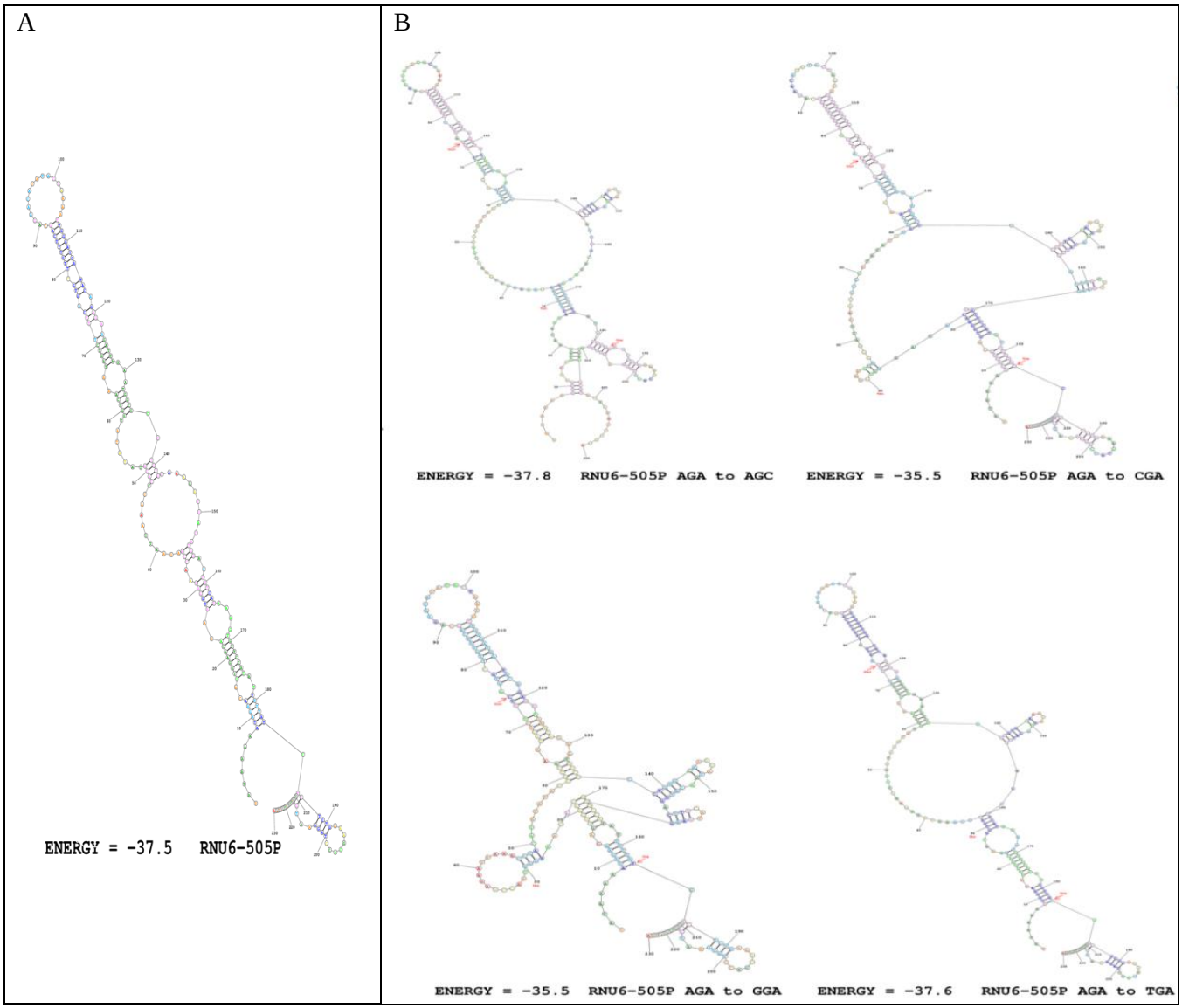

Figure 12: *Secondary structure prediction of RNU6-505P with (B) and without (A) 1-nucleotide mutation on AGA*

3.5 Mutation Prediction of AGA is Fatal for the Structure of RNU6-505P

What happens to the secondary structure if we consider all possible one-nucleotide mutation on that regions? The next picture shows the prediction of the secondary structure of RNU6-505P *"extended"* sequence (345bp) obtained by RNAfold program, while the next one depicts the predictions of the secondary structures obtained mutating the AGA sequence. Amazingly, taking a look to the reported predictions (Figure 12), it is possible to deduce that mutations on AGA nucleotides could be fatal for the RNA structure of the sequence. Evidence that sustains our hypothesis that RNU6-505P, even if is *not* a protein coding sequence, could have a relevant rule in the cerebral development and evolution from the apes.

4 Conclusions

Pseudogenes have long been considered mere remnants of evolution. Only lastly, scientific research tends to recognize an important regulatory function in the genetic context. This work has highlighted how the pseudogene RNU6-505P, while not coding for any protein, is a potential marker for schizophrenic patients. In addition, clustering analysis has shown that related genes are involved in neurological functions and pathologies, advancing the hypothesis that RNU6-505P is engaged in some prenatal development of the human brain.

Finally, the evolutionary pressure to which it has been subjected and the prediction of the second-level structure in the case of mutation on a single nucleotide, have further emphasized the importance that this pseudogene may have had in the human evolutionary process.

5 Declarations

5.1 Limitations of the Study

In our opinion here, we present the first case of highlighting role of RNU6-505P pseudogene in schizophrenia pathogenesis. The study has some limitations that can be overcome in future studies: 1) low number of schizophrenia cases and control samples studied in transcriptome analyses; 2) "mouse model" *in vitro* analysis needed to confirm biological significance of our findings on correlation of above mentioned up- and down-regulated genes expression. There are also a number of the most recent studies utilizing novel approaches appears showing newly discovered candidate genes, risk genes, and marker genes for schizophrenia [29-32].

5.2 Acknowledgments

Acknowledgement to dr. Liu's lab and all of our colleagues that daily support our research. We appreciate the help and the big patience our families given us. We thank the 3 anonymous reviewers for their so-called insights.

5.3 Funding Source

This work was supported by grants of National Natural Science Foundation of China (NSFC81671980, 31700126, 81871623, 81971910). The funding bodies played no roles in the design of the study; the collection, analysis or interpretation of data; or in writing the manuscript.

5.4 Competing Interests

The author declared that no conflict of interest exists in this publication.

How to Cite this Article:

E. Mastriani, A. Rakov, and S.-L. Liu, "Differential Expression Analysis on Schizophrenia Dataset Suggests Pseudogene RNU6-505P as under Selective Pressure", *Int. Ann. Sci.*, vol. 9, no. 1, pp. 86-99, Jan. 2020. doi: 10.21467/ias.9.1.86-99

References

[1] O. de Manzano, S. Cervenka, A. Karabanov, L. Farde, and F. Ullen, "Thinking outside a less intact box: thalamic dopamine D2 receptor densities are negatively related to psychometric creativity in healthy individuals," *PLoS One*, vol. 5, no. 5, p. e10670, May 17 2010.

[2] I. R. Reddy, J. Ukrani, V. Indla, and V. Ukrani, "Creativity and psychopathology: Two sides of the same coin?," *Indian J Psychiatry*, vol. 60, no. 2, pp. 168-174, Apr-Jun 2018.

[3] J. Parnas, K. E. Sandsten, C. H. Vestergaard, and J. Nordgaard, "Schizophrenia and Bipolar Illness in the Relatives of University Scientists: An Epidemiological Report on the Creativity-Psychopathology Relationship," *Front Psychiatry*, vol. 10, p. 175, 2019.

[4] C. Asanuma, W. T. Thach, and E. G. Jones, "Cytoarchitectonic delineation of the ventral lateral thalamic region in the monkey," *Brain Res*, vol. 286, no. 3, pp. 219-35, May 1983.

[5] S. Sabunciyan, "Gene Expression Profiles Associated with Brain Aging are Altered in Schizophrenia," *Sci Rep*, vol. 9, no. 1, p. 5896, Apr 11 2019.

[6] M. Li *et al.*, "A powerful conditional gene-based association approach implicated functionally important genes for schizophrenia," *Bioinformatics*, vol. 35, no. 4, pp. 628-635, Feb 15 2019.

[7] M. Manchia *et al.*, "Pattern of gene expression in different stages of schizophrenia: Down-regulation of NPTX2 gene revealed by a meta-analysis of microarray datasets," *Eur Neuropsychopharmacol*, vol. 27, no. 10, pp. 1054-1063, Oct 2017.

[8] O. Bergman, R. Karry, J. Milhem, and D. Ben-Shachar, "NDUFV2 pseudogene (NDUFV2P1) contributes to mitochondrial complex I deficits in schizophrenia," *Mol Psychiatry*, Dec 10 2018.

[9] K. J. Brennand *et al.*, "Modelling schizophrenia using human induced pluripotent stem cells," *Nature*, vol. 473, no. 7346, pp. 221-5, 2011.

[10] B. S. Carvalho and R. A. Irizarry, "A framework for oligonucleotide microarray preprocessing," *Bioinformatics*, vol. 26, no. 19, pp. 2363-7, Oct 1 2010.

[11] R. C. Gentleman *et al.*, "Bioconductor: open software development for computational biology and bioinformatics," *Genome Biol*, vol. 5, no. 10, p. R80, 2004.

[12] M. E. Ritchie *et al.*, "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Res*, vol. 43, no. 7, p. e47, Apr 20 2015.

[13] L. Scrucca, M. Fop, T. B. Murphy, and A. E. Raftery, "mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models," *R J*, vol. 8, no. 1, pp. 289-317, Aug 2016.

[14] U. K. Muppirala, V. G. Honavar, and D. Dobbs, "Predicting RNA-protein interactions using only sequence information," *BMC Bioinformatics*, vol. 12, p. 489, Dec 22 2011.

[15] B. A. Lewis *et al.*, "PRIDB: a Protein-RNA interface database," *Nucleic Acids Res*, vol. 39, no. Database issue, pp. D277-82, Jan 2011.

[16] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *J Mol Biol*, vol. 215, no. 3, pp. 403-10, Oct 5 1990.

[17] G. Feng, P. Shaw, S. T. Rosen, S. M. Lin, and W. A. Kibbe, "Using the bioconductor GeneAnswers package to interpret gene lists," *Methods Mol Biol*, vol. 802, pp. 101-12, 2012.

[18] J. Paquette and T. Tokuyasu, "EGAN: exploratory gene association networks," *Bioinformatics*, vol. 26, no. 2, pp. 285-6, Jan 15 2010.

[19] D. Maglott, J. Ostell, K. D. Pruitt, and T. Tatusova, "Entrez Gene: gene-centered information at NCBI," *Nucleic Acids Res*, vol. 39, no. Database issue, pp. D52-7, Jan 2011.

[20] J. S. Amberger, C. A. Bocchini, F. Schiettecatte, A. F. Scott, and A. Hamosh, "OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders," *Nucleic Acids Res*, vol. 43, no. Database issue, pp. D789-98, Jan 2015.

[21] S. Lindsay and A. J. Copp, "MRC-Wellcome Trust Human Developmental Biology Resource: enabling studies of human developmental gene expression," *Trends Genet*, vol. 21, no. 11, pp. 586-90, Nov 2005.

[22] J. Huerta-Cepas, F. Serra, and P. Bork, "ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data," *Mol Biol Evol*, vol. 33, no. 6, pp. 1635-8, Jun 2016.

[23] Z. Yang, "PAML 4: phylogenetic analysis by maximum likelihood," *Mol Biol Evol*, vol. 24, no. 8, pp. 1586-91, Aug 2007.

[24] J. P. Bielawski and Z. Yang, "A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution," *J Mol Evol*, vol. 59, no. 1, pp. 121-32, Jul 2004.

[25] Z. Yang and R. Nielsen, "Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages," *Mol Biol Evol*, vol. 19, no. 6, pp. 908-17, Jun 2002.

[26] J. Zhang, R. Nielsen, and Z. Yang, "Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level," *Mol Biol Evol*, vol. 22, no. 12, pp. 2472-9, Dec 2005.

[27] R. B. Denman, "Using RNAFOLD to predict the activity of small catalytic RNAs," *Biotechniques*, vol. 15, no. 6, pp. 1090-5, Dec 1993.

[28] G. Stelzer *et al.*, "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses," *Curr Protoc Bioinformatics*, vol. 54, pp. 1 30 1-1 30 33, Jun 20 2016.

[29] Q. Wang *et al.*, "A Bayesian framework that integrates multi-omics data and gene networks predicts risk genes from schizophrenia GWAS data," *Nat Neurosci*, vol. 22, no. 5, pp. 691-699, May 2019.

[30] Q. Xie, W. Shen, Z. Li, A. Baranova, H. Cao, and Z. Li, "A core collection of pan-schizophrenia genes allows building cohort-specific signatures of affected brain," *Sci Rep*, vol. 9, no. 1, p. 12671, Sep 3 2019.

[31] N. E. Clifton *et al.*, "Dynamic expression of genes associated with schizophrenia and bipolar disorder across development," *Transl Psychiatry*, vol. 9, no. 1, p. 74, Feb 4 2019.

[32] V. K. Ota *et al.*, "Gene expression over the course of schizophrenia: from clinical high-risk for psychosis to chronic stages," *NPJ Schizophr*, vol. 5, no. 1, p. 5, Mar 28 2019.