Genomics & Informatics Vol. 6(1) 32-35, March 2008

Functional Prediction of Imprinted Genes in Chicken Based on a Mammalian Comparative Expression Network

HyoYoung Kim[1], Sunjin Moon[1] and Heebal Kim[1,2]*

[1]Department of Agricultural Biotechnology, Seoul National University, Seoul 151-742, Korea, [2]Graduate Program in Bioinformatics, Seoul National University, Seoul 151-742, Korea

Abstract

Little evidence supports the existence of imprinted genes in chicken. Imprinted genes are thought to be intimately connected with the acquisition of parental resources in mammals; thus, the predicted lack of this type of gene in chicken is not surprising, given that they leave their offspring to their own heritance after conception. In this study, we identified several imprinted genes and their orthologs in human, mouse, and zebrafish, including 30 previously identified human and mouse imprinted genes. Next, using the HomoloGene database, we identified six orthologous genes in human, mouse, and chicken; however, no orthologs were identified for *SLC22A18*, and mouse *Ppp1r9a* was not included in the HomoloGene database. Thus, from our analysis, four candidate chicken imprinted genes *(IGF2, UBE3A, PHLDA2,* and *GRB10)* were identified. To expand our analysis, zebrafish was included, but no probe ID for UBE3A exists in this species. Thus, ultimately, three candidate imprinted genes *(IGF2, PHLDA2,* and *GRB10)* in chicken were identified. *GRB10* was not significant in chicken and zebrafish based on the Wilcoxon-Mann-Whitney test, whereas a weak correlation between *PHLDA2* in chicken and human was identified from the Spearman's rank correlation coefficient. Significant associations between human, mouse, chicken, and zebrafish were found for *IGF2* and *GRB10* using the Friedman's test. Based on our results, *IGF2, PHLDA2*, and *GRB10* are candidate imprinted genes in chicken. Importantly, the strongest candidate was *PHLDA2*.

Keywords: chicken, conservation, homologs, imprinted genes, statistical analysis

Introduction

Imprinted genes are not inherited in a recessive or dominant fashion (http://www.hopkinsmedicine.org/press/2002/November/epigenetics.htm); instead, they are monoallelic, meaning that they are epigenetically expressed from a single parent-specific allele (either paternal [sperm] or maternal [egg]). Such genes are also asynchronously replicated from pre-imprinted chromosomes (Reik and Walter., 2001). Imprinting is believed to be important in placental mammals, because it may affect the transfer of resources between mother and offspring; however, imprinted genes also exist in higher seed plants, which utilize a placenta-like tissue known as endosperm to nourish the developing embryo. Even egg-laying mammals (i.e., monotremes) show imprinting in suckling-related genes. However, the existence of imprinted genes in chicken *(Gallus gallus)* is controversial (Miguel *et al*., 2004). For example, *IGF2*, which is paternally expressed in marsupials (e.g., possums) and mammals, is not similarly expressed in birds (Yokomine *et al*., 2005). Nonetheless, the arrangement and substance of the chicken genome is highly conserved in many human imprinted domains, including the human imprinted gene cluster that contains *IGF2, H19, KCNQ1, ASCL2,* and *CDKN1C* (Rapkins *et al*., 2006). If, as has been suggested, imprinted genes are intimately connected with the acquisition of parental resources, we would not anticipate the existence of such genes in chicken, which leave their offspring to their own heritance after conception. Phylogenetic analyses expose that the relationship between human and mouse is closer than that between human, mouse, and chicken. Similarly, the relationship between zebrafish and chicken is quite distant (Shah *et al*., 2004). Nonetheless, we assumed that chicken have imprinted genes due to the existence of common ancestral genomic regions that have evolved on a similar basis in each of the aforementioned species. The purpose of this study was to identify candidate imprinted genes in chicken based on an analysis of orthologous genes in human, mouse, zebrafish, and chicken using the HomoloGene database.

*Corresponding author: E-mail heebal@snu.ac.kr
Tel +82-2-880-4803, Fax +82-2-883-8812
Accepted 22 January 2008

Methods

Data selection

Human, mouse, chicken, and zebrafish were selected as our experimental units. All gene expression data for these species were compiled into a CEL file using the GEO (Gene Expression Omnibus) database at the National Center for Biotechnology Information (NCBI). A list of imprinted human and mouse genes were obtained from http://www.geneimprint.com/site/genes-by-species; the probe ID and name of each gene were downloaded from http://www.affymetrix.com/support/technical/annotationfilesmain.affx to confirm the information. All data were subsequently compiled in HomoloGene (ftp://ftp.ncbi.nih.gov/pub/HomoloGene/FTPsite/build57).

Statistical analysis

To determine how strongly the various data were connected, we calculated correlations for all of the genes in the network using a hard cutoff, with 1 signifying an absolute correlation for those values greater than 0.5 and 0 for all other values. We then summed the re-coded values to analyze connectivity strength. The data were also analyzed using nonparametric statistical tests. To test for differences between each pair of species, we used the binomial exact test and the Wilcoxon-Mann-Whitney test. We also computed Spearman's and Kendall's correlation coefficients to analyze the relationships between pairs of species, and the Kolmogorov-Smirnov test was used to confirm differences between pairs of species. Finally, the Friedman's test was used to identify differences between more than three species. All statistical tests used in this study were performed using Python and statistical package R (R-project, http://www.r-project.org/).

Results and Discussion

Identification of orthologous genes in human, mouse, chicken, and zebrafish using HomoloGene

Given that the imprinted genes in human and mouse are known, we selected them as our experimental units. The imprinted status of each species was downloaded from the Internet (http://www.geneimprint.com/site/genes-by-species). Thirty orthologous imprinted genes were found in human and mouse. We next used the HomoloGene database to search for homologous genes in human, mouse, and chicken. Of the 24, 17, and seven genes identified in the three species, respectively, six were found to be orthologous (*PPP1R9A, IGF2, SLC22A18, PHLDA2, UBE3A,* and *GRB10*; Table 1).

Identification of candidate imprinted genes in chicken

We calculated correlation values for each of the genes using a hard cutoff, and then summed the recoded connection strengths, as shown in Table 1 (see the Materials and Methods). From this result, we confirmed five highly conserved orthologous genes *(IGF2, SLC22A18, PHLDA2, UBE3A,* and *GRB10)* in human, mouse, and chicken. This represents the average connectivity for identical genes with different probe IDs. We next computed connectivity values for the genes in each species, and found that *SLC22A18* was not orthologous between human, mouse, and chicken. Consequently, using the binomial exact test, we identified four potential imprinted genes in chicken *(IGF2, PHLDA2, UBE3A,* and *GRB10)*.

Statistical analysis

In addition to the above analysis, we conducted a series

Table 1. Average connectivity defined by Homologene for each species

Gene	Human		Mouse		Gallus	
	Connectivity*	Sequence†	Connectivity	Sequence	Connectivity	Sequence
IGF2 (P)	1,515.7	NP_000603.1	885.5	NP_034644.1	3,357.5	XP_421026.2
SLC22A18 (M)	28.0	NP_899056.1	1,564.0	NP_032793.1	674.0	XP_421021.2
PHLDA2 (M)	1,324.0	NP_003302.1	2,578.0	NP_033460.1	1,649.5	XP_421020.1
UBE3A (M)	1,358.3	NP_570854.1	376.0	NP_035798.2	5,515.0	XP_416882.1
GRB10	73.0	NP_005302.3	359.7	NP_034475.2	1,428.5	XP_001034371.1

*How strongly the various data do what?

†Reference sequence ID

P, paternal; M, maternal.

Table 2. Statistical analysis to identify differences in correlation connectivity between species

Species	IGF2		PHLDA2		GRB10	
	p value*	r†	p value	r	p value	r
G-H	<2.2e-16	0.0078	<2.2e-16	0.1953	2.29E-05	0.1496
G-M	<2.2e-16	−0.1444	6.66E-16	−0.0394	5.09E-11	−0.0992
G-Z	6.06E-11	0.0188	3.57E-07	−0.1670	0.2451	−0.1170
H-M	3.08E-05	−0.0523	1.33E-10	0.1083	0.08541	−0.0096
M-Z	3.93E-12	−0.0095	2.81E-09	−0.2053	<2.2e-16	0.3644
Z-H	<2.2e-16	0.2743	9.30E-14	0.0493	5.96E-08	0.4231

*To test for differences between each pair of species using Wilcoxon-Mann-Whitney test
†To analyze the relationships between pairs of species using Spearman correlation
G, chicken; H, human; M, mouse; Z, zebrafish; G-H represents the association between chicken and human.; G-M represents the association between chicken and mouse.; G-Z represents the association between chicken and zebrafish.

of nonparametric tests with zebrafish added to the list of species. No probe ID was identified for *UBE3A* in zebrafish, so we excluded it from our analysis. To test for differences in distribution between each pair of species, we used the Wilcoxon-Mann-Whitney test and the Kolmogorov-Smirnov test. According to our results, *GRB10* was different chicken and human. We also calculated Spearman's and Kendall's correlation coefficients to compare the relationships between each pair of species. A weak relationship was identified for *PHLDA2* between chicken and human, which reached a significance level of α=0.05 (Table 2). Finally, we used the Friedman's test to compare differences in association among the four species. *IGF2* was not significant in any case, while *GRB10* was not significant in human, mouse, and chicken (Table 3).

Table 3. Statistical analysis to identify differences in connectivity between species using Friedman test

Species	IGF2	PHLDA2	GRB10
H_M_G_Z	0.0719	0.0169	NA
H_M_G	0.0970	0.0388	0.2231
H_M_Z	0.0970	0.0388	0.0907
M_G_Z	0.0970	0.0388	0.0183

G, chicken; H, human; M, mouse; Z, zebrafish; H-M- G-Z represents the association between human, mouse, chicken, and zebrafish.; H-M-G represents the association between human, mouse and chicken.; M-G-Z represents the association between mouse, chicken, and zebrafish.; NA stands for not available.

Comparison of our data with comparative data

To compare our data, referred to as comparative data, we computed correlations for all data and summed the values. We considered only those genes that were related to chicken. Based on the Wilcoxon-Mann-Whitney test and the Kolmogorov-Smirnov test, *GRB10* was different chicken and zebrafish using our data; however, it was significant using comparative data. In contrast, *PHLDA2* was different chicken and mouse using our data, while it was not significant using comparative data. Using Spearman's and Kendall's correlations, *PHLDA2* was shown to be weakly related between chicken and human using our data; however, no correlation was found using comparative data. Finally, a weak correlation was identified for *GRB10* using comparative data, but not using our data. The Friedman's test for *IGF2* produced identical results regardless of whether our data or comparative data were used, but for *PHLDA2*, the result of comparative data was not the same for human-mouse-chicken and mouse-chicken-zebrafish using our data. Thus, *IGF2, PHLDA2*, and *GRB10* were identified as putative imprinted genes in chicken. Importantly, the strongest candidate was *PHLDA2*. As more genomic data become available, we plan to repeat our analysis to identify additional candidate imprinted genes in chicken.

Acknowledgements

This work was supported by a grantfrom BioGreen 21 Program (20050301034487), Rural Development Administration, Republic of Korea.

References

Chung, M.K. (2005). Possible gene expression and imprinting defects in assisted reproductive technology. *The 14th Korea Genome Conference* pp.50.

Colosi, D.C., Martin, D., More`, K., and Lalande, M. (2006). Genomic organization and allelic expression of UBE3A in chicken. *Gen*. 383, 93-98.

Crawley, M.J. (2005). Statistics an introduction using R (UK: Imperial College London), pp.80-100.

Dunzinger, U., Haaf, T., and Zechner, U. (2007). Conserved synteny of mammalian imprinted genes in chicken, frog, and fish genomes. *Cytogent. Genome Res.* 117, 78-85.

Khatib, H., Zaitoun, I., and Kim, E.S. (2007). Comparative analysis of sequence characteristics of imprinted genes in human, mouse and cattle. *Mamm. Genome*. 18, 538-547.

McKillup, S. (2006). Statistics explained (The United Kingdom: Cambridge University Press), pp.225-243.

Miguel, C., Kelsey, G., and Reik, W. (2004). Resourceful imprinting. *Nature* 432, 53-57.

Oldham, M.C., Horvath, S., and Geschwind, D.H. (2006). Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *PNAS*. 103, 17973-17978.

Park, M.H., Ku, H.J., Lee, H.J., Kim, K.J., Park, C., Oh, B.S., Kimm, K.C., and Lee, J.Y. (2005). Single nucleotide polymorphisms of a 16kb region on human chromosome 11 p15.5 that includes the h19 gene. *Genomics and Informatics* 3, 74-79.

Rapkins, R.W., Hore, T., Smithwick, M., Ager, E., Pask, A.J., Renfree, M.B., Kohn, M., Hameister, H., Nicholls, R.D., Deakin, J.E., and Graves, J.A. (2006). Recent assembly of an imprinted domain from non-imprinted components. *PLoS Genet*. 2, e182.

Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. *Nat. Rev. Gnet*. 2, 21-32.

Shah, N., Couronne, O., Pennacchio, L.A., Bruno, M., Batzoglou, S., Bethel, E.W., Rubin, E.M., Hamann. B., and Dubchak, I. (2004). Phylo-VISTA: an interactive visualization tool for multiple dna sequence alignments. *Bioinformatics* 20, 636-643.

Yokomine, T., Shirohzu, H., Purbowasito, W., Toyoda, A., Iwama, H., Ikeo, K., Hori, T., Mizuno, S., Tsudzuki, M., Matsuda, Y.I., Hattori, M., Sakaki, Y., and Sasaki, H. (2005). Structural and functional analysis of a 0.5-Mb chicken region orthologous to the imprinted mammalian Asc12/Mash2-Igf2-H19 region. *Genome Research* 15, 154-165.

Zaitoun, I., and Khatib, H. (2007). Comparative genomics imprinting and expression analysis of six cattle genes. Animal Science. *http://www.hopkinsmedicine.org/press/2002/November/epigenetics.htm.*