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

HyoYoung Kim1, Sunjin Moon1 and Heebal Kim1,2*

1Department of Agricultural Biotechnology, Seoul National University, Seoul 151-742, Korea, 2Graduate 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.
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 recorded 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

<table>
<thead>
<tr>
<th>Gene</th>
<th>Human</th>
<th>Mouse</th>
<th>Gallus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Connectivity*</td>
<td>Sequence†</td>
<td>Connectivity</td>
</tr>
<tr>
<td>IGF2 (P)</td>
<td>1,515.7</td>
<td>NP_000603.1</td>
<td>885.5</td>
</tr>
<tr>
<td>SLC22A18 (M)</td>
<td>28.0</td>
<td>NP_089095.1</td>
<td>1,564.0</td>
</tr>
<tr>
<td>PHLDA2 (M)</td>
<td>1,324.0</td>
<td>NP_003302.1</td>
<td>2,578.0</td>
</tr>
<tr>
<td>UBE3A (M)</td>
<td>1,358.3</td>
<td>NP_507845.1</td>
<td>376.0</td>
</tr>
<tr>
<td>GRB10</td>
<td>73.0</td>
<td>NP_005302.3</td>
<td>359.7</td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th>Species</th>
<th>IGF2</th>
<th>PHLDA2</th>
<th>GRB10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p value*</td>
<td>r †</td>
<td>p value</td>
</tr>
<tr>
<td>G-H</td>
<td>&lt;2.2e-16</td>
<td>0.0078</td>
<td>&lt;2.2e-16</td>
</tr>
<tr>
<td>G-M</td>
<td>&lt;2.2e-16</td>
<td>-0.1444</td>
<td>6.66E-16</td>
</tr>
<tr>
<td>G-Z</td>
<td>6.06E-11</td>
<td>0.0188</td>
<td>3.57E-07</td>
</tr>
<tr>
<td>H-M</td>
<td>3.08E-05</td>
<td>-0.0523</td>
<td>1.33E-10</td>
</tr>
<tr>
<td>M-Z</td>
<td>3.93E-12</td>
<td>-0.0095</td>
<td>2.81E-09</td>
</tr>
<tr>
<td>Z-H</td>
<td>&lt;2.2e-16</td>
<td>0.2743</td>
<td>9.30E-14</td>
</tr>
</tbody>
</table>

*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.

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

<table>
<thead>
<tr>
<th>Species</th>
<th>IGF2</th>
<th>PHLDA2</th>
<th>GRB10</th>
</tr>
</thead>
<tbody>
<tr>
<td>H,M,G,Z</td>
<td>0.0719</td>
<td>0.0169</td>
<td>NA</td>
</tr>
<tr>
<td>H,M,G</td>
<td>0.0970</td>
<td>0.0388</td>
<td>0.2231</td>
</tr>
<tr>
<td>H,M,Z</td>
<td>0.0970</td>
<td>0.0388</td>
<td>0.0607</td>
</tr>
<tr>
<td>M,G,Z</td>
<td>0.0970</td>
<td>0.0388</td>
<td>0.0183</td>
</tr>
</tbody>
</table>

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 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 $\alpha = 0.05$ (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).

References


