Adult Tissue-specific Expression of a Dppa3-derived Retrongene Represents a Postnatal Transcript of Pluripotent Cell Origin

Received for publication, October 14, 2005
Published, JBC Papers in Press, November 15, 2005, DOI 10.1074/jbc.C500415200

Stephen J. Elliman 1, Isaac Wu 1†, and Daniel M. Kemp 1‡

From the 1Diabetes and Metabolism Disease Area, Novartis Institutes for BioMedical Research, Cambridge, Massachusetts 02139 and the 2Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

Processed pseudogenes emerge by reverse transcription of spliced mRNAs followed by incorporation of the resultant cDNA into the genome. Their genesis requires that retrotransposition occurs within the germ line, a provision that significantly limits random distribution of source genes. We previously identified embryonic stem cell-specific genes as an enriched source of retropseudogene origin. Nanog, Oct4, and Dppa3 (Stella/PGC7) presented as source genes for > 30 processed pseudogenes within the human genome. In the current study, we extended our previous analysis and focused on the pluripotent cell-specific Dppa3 gene family. Of the five Dppa3 genes characterized, four were associated with putative retropseudogenes as determined by nucleotide BLAST (basic local alignment sequence tool) searches of the respective mRNA transcripts against the human genome. A subset of the 11 Dppa3-derived hits were then screened against a human adult tissue cDNA panel for evidence of transcriptional activity. One of the putative Dppa3-derived retropseudogenes, Dppa3(d), located on human chromosome 16p13, tested positive for mRNA transcript in bone marrow, peripheral blood, pancreas, adrenal gland, and thyroid gland. Specificity against the source Dppa3 gene expression was sequenced verified, and independent human tissue samples were obtained to confirm Dppa3(d) expression. These data substantiate the existence of human adult tissue-specific transcripts that originate via retrotransposition of the pluripotent cell-specific gene, Dppa3. Further studies may reveal an evolutionary role for this example of genetic diversity, but in the short term our observations serve a cautionary purpose regarding the use of Dppa3 transcripts in adult tissue-derived cells as a potential marker of pluripotency.

Processed pseudogenes (retropseudogenes) emerge by reverse transcription of processed mRNA and consequent genomic integration (1–3). Given the absolute requirement for germ line transmission to realize a processed pseudogene, we hypothesized that pluripotent tissue-specific gene transcripts may possess a competitive advantage over non-germ cell-specific genes with respect to inherited retrotransposition. Previously, we reported the excessive frequency and apparent selectivity of retrotransposition of pluripotent cell-specific genes. Six highly homologous pseudogenes for Oct4, 10 for Nanog, and 16 for Dppa3 (Stella/Pgc7) were identified by nucleotide BLAST (basic local alignment sequence tool) searches against the respective gene mRNA transcripts. Of 15 non-ES cell-specific transcription factor genes, only one had a single pseudogene hit in our screen, emphasizing the apparent selectivity (4). Oct4 and Nanog are both required to maintain pluripotency of embryonic cell-types (5–7), whereas the functional role of Dppa3 remains unclear. Studies suggest its function may be related to the cleavage stages of embryogenesis, rather than in germ cell specification (8–12). The nuclear localization of Dppa3 within the inner cell mass of blastocysts suggests a critical function, maybe as a transcriptional regulator (9).

Although the majority of retrotranspositions appear to manifest as nonfunctional genetic artifact, exceptions exist whereby retrogenes exhibit transcriptional activity (13, 14) and encode functional proteins (15, 16). The evolutionary implications of functional retrogenes are well defined (17), but the role of the embryonic cell transcriptome in this context remains to be established. Here, we extended our previous investigation to analyze additional pluripotent cell-specific genes for their predisposition to retrotransposition. In addition, we screened a panel of 24 adult human tissues to reveal transcriptional activity of these novel processed retrogenes. We report that one of the Dppa3-derived genomic sequences is transcriptionally active and displays a tissue-specific expression profile.

EXPERIMENTAL PROCEDURES

BLAST was used to compare the cDNA of Nanog, Oct4, Dppa3 (Stella/PGC7), GDF3, and Dppa1, -2, -4, and -5 with the National Center for Biotechnology Information human genome data base (18). An E value limit was set to $1 \times 10^{-3}$ for these searches, and the Ensembl genome browser was used to confirm all blast hits. Processed pseudogenes were selected based on the search criteria, which calculated similarity to the source genes in question. Pseudogene sequences were compared with their corresponding source gene sequence and variations were exploited to design specific primers to each retropseudogene using Vector NTI Suite 9. All of the primers associated with one source gene were cross-checked with its source gene and all of its sister pseudogenes to confirm specificity. Primers were purchased from Sigma Genosys and resuspended to 1 μg/μl in ddH2O and then diluted 200-fold in ddH2O. PCR against the human genome was performed using PCR SuperMix (Invitrogen).

Human rapid scan panels were purchased from OriGene Technologies. The panels contained 96 wells consisting of 24 tissues arrayed in quadruplicate. Each quadruplicate consisted of increasing concentrations of cDNA (1×, 10×, 100×, and 1000×). PCR using source- and retropseudogene-specific primers along with PCR SuperMix was carried out against these panels. In all panels screened in this study, β-actin was amplified to validate the integrity of the cDNA samples.

RNA was prepared from donor-derived human blood and bone marrow (Cambrex) using RNeasy kits (Qiagen). First-strand cDNA synthesis was performed using the SuperScript III RT kit (Invitrogen). PCR was carried out using source and retropseudogene primers as well as PCR SuperMix. β-Actin was employed as a control to validate the integrity of each sample.

Pancreas cDNA was obtained from BioChain, and RACE (rapid amplification of cDNA ends) cDNA was purchased from Ambion. cDNAs isolated from thyroid gland, heart and adrenal gland were obtained from BioChain.

Each reaction well contained 5 μl of forward primer, 5 μl of reverse primer, 1 μl of genomic DNA, and 39 μl of PCR SuperMix (Invitrogen). These samples were run at 94 °C (5 min), then through 35 cycles of the following three steps (1) 94 °C (1 min), (2) 55 °C (1 min), (3) 72 °C (1 min), and finally at 72 °C (5 min) followed by refrigeration. Annealing temperatures of 50 and 60 °C were also attempted for each primer pair. PCR products were then run on agarose gels, treated with ethidium bromide, and imaged. For the rapid scan screens, 5 μl of forward primer, 5 μl of reverse primer, and 39 μl of PCR SuperMix (Invitrogen) was added to each well on the cDNA panels. PCR and electrophoresis were carried out as detailed above.

RESULTS

Previously, we BLASTed four known pluripotent cell-specific mRNA transcripts against the human genome and revealed their predisposition for retrotransposition and subsequent generation of processed pseudogenes (4). The frequency of retrotransposition within the “pseudogene” set was striking, compared with that of non-pluripotent cell-specific genes. In summary, three of the four stem cell-specific genes contributed to a total of 32 processed pseudogenes, whereas 15 non-stem cell-specific genes contributed to a single processed pseudogene.
In the current study, we shifted focus to a different set of recently identified pluripotent tissue-specific genes. The Dppa family of genes (Dppa1–5) are coordinately expressed in pluripotent tissues, including blastocyst, epiblast, and primordial germ cells (11). We observed by BLAST analysis that four of the five Dppa genes possess multiple, highly conserved retropseudogenes, dispersed throughout the human genome (Table 1). Briefly, mRNA transcript sequences of these genes were individually BLASTed against the National Center for Biotechnology Information human genome data base. The hits that represented the source genes contained the expected intron/exon gene organization, but all other sequences were contiguous, such that the conserved alignments reflected the mRNA transcript sequence rather than the genomic sequence. This suggested that these conserved sequences likely resulted from retrotransposition events (19). In the instances of Dppa1 and Dppa5, the mouse ortholog was used as source gene in the absence of an annotated human variant. However, these murine transcripts were BLASTed against the human genome. MEGABLAST was employed with an expected value significance cut-off equal to 0.01, and this relatively high stringency was used to predict the occurrence of gene retrotransposition. Due to this relative increase in stringency compared with our previous study (19), and the relatively high stringency was used to predict the occurrence of gene retrotransposition. This led to the identification of 20 gene-related retropseudogenes, the mouse ortholog of which matched the human gene with a high identity. The mouse orthologs of the Dppa1–5 genes were associated with 22 processed pseudogenes by this bioinformatic analysis. In summary, these data further consolidate our previous observation of apparent predisposition of pluripotent cell-specific genes to the generation of processed pseudogenes (38 from 8 ES cell-specific source genes, see Table 1).

Dppa3 (Stella/Pgc7) was associated with 11 processed pseudogenes, annotated Dppa3 (a–k) (Table 2). A non-extensive bioinformatic analysis was performed on each putative pseudogene to predict the likelihood of transcriptional activity. Start and stop codons conserved from the source gene represented a potential identifier, as did conservation of the full coding region (Table 2). Following this analysis, four processed pseudogenes of Dppa3 were selected to further characterize their functional potential (i.e. Dppa3 (a–d)). To selectively identify processed pseudogenes of Dppa3, specific PCR primers were designed to mismatch with other pseudogenes as well as the source gene (Table 3). All primer pairs generated a single product of the predicted size, and sequencing confirmed specificity (data not shown).

To determine whether any processed pseudogenes of Dppa3 were transcriptionally active, we screened each pseudogene-specific PCR primer pair against a panel of 24 adult tissue-specific cDNA pools. Notably, the Dppa3 source gene primer pair failed to generate an amplification product in any of the adult tissues tested (data not shown), confirming the reported specificity of Dppa3 expression in pluripotent tissues (11). In contrast, RT-PCR against Dppa3(d) generated amplification products of the predicted size. Transcripts for Dppa3(d) were detected in bone marrow, peripheral blood, adrenal gland, pancreas, and thyroid (thyroid not clearly seen in Fig. 1). In each case, the amplified product was only detected in the 1000x cDNA reaction, due to the low expression level of the transcript in all positive tissue types.

To validate the expression of Dppa3(d) in these specific tissues, we obtained human RNA from alternative sources, from which to perform RT-PCR with the putative pseudogene-selective primers. Human bone marrow-derived RNA was extracted from mononuclear cells that were isolated from donated whole bone marrow aspirates. Peripheral blood was similarly donated by volunteers, from which mononuclear cells were isolated and prepared for RNA extraction. Other human tissues were acquired from various vendors (see Materials and Methods) and RT-PCR was performed on all samples simultaneously (Fig. 2). Amplification of the Dppa3(d) product was successful in all of the re-tested adult tissue-types, confirming the data from the primary tissue panel screens. Importantly, the Dppa3 source gene primers were unable to amplify any product within these tissue samples, consistent with the reported absence of Dppa3 source gene transcript in human adult tissues (Fig. 2). The variation in amplification product intensity within this panel of tissues may be explained by the variation in sample source, cell-type homogeneity, as well as the non-quantitative nature of gel-based RT-PCR.

To confirm that the Dppa3(d) primers were not selective for the source Dppa3 gene-derived transcript, and that the primers for the source gene were legitimate, RT-PCR was performed on NT2 cell-derived RNA (a teratocarcinoma cell line). This cell line expresses Nanog and Oct4, and differentiates preferentially to neuronal lineages upon retinoic acid treatment (20). The Dppa3(d) primers failed to generate any product from NT2 cell RNA, whereas in bone marrow and pancreas (obtained from alternative sources, see “Experimental Procedures”), robust transcripts were amplified under the same PCR conditions (Fig. 3A). In contrast, the Dppa3 source gene primer pair successfully amplified a product of the expected size from NT2 cell RNA, confirming expression of Dppa3 coordinate with the pluripotency-related genes Nanog and Oct4 (Fig. 3B). However, no amplification was observed in bone marrow and pancreas. To determine whether the source gene primers were able to amplify any product of the processed pseudogene Dppa3(d) transcript, we performed RT-PCR with the putative pseudogene-selective primers. Human bone marrow-derived RNA was extracted from mononuclear cells that were isolated from donated whole bone marrow aspirates. Peripheral blood was similarly donated by volunteers, from which mononuclear cells were isolated and prepared for RNA extraction. Other human tissues were acquired from various vendors (see Materials and Methods) and RT-PCR was performed on all samples simultaneously (Fig. 2). Amplification of the Dppa3(d) product was successful in all of the re-tested adult tissue-types, confirming the data from the primary tissue panel screens. Importantly, the Dppa3 source gene primers were unable to amplify any product within these tissue samples, consistent with the reported absence of Dppa3 source gene transcript in human adult tissues (Fig. 2). The variation in amplification product intensity within this panel of tissues may be explained by the variation in sample source, cell-type homogeneity, as well as the non-quantitative nature of gel-based RT-PCR.

To confirm that the Dppa3(d) primers were not selective for the source Dppa3 gene-derived transcript, and that the primers for the source gene were legitimate, RT-PCR was performed on NT2 cell-derived RNA (a teratocarcinoma cell line). This cell line expresses Nanog and Oct4, and differentiates preferentially to neuronal lineages upon retinoic acid treatment (20). The Dppa3(d) primers failed to generate any product from NT2 cell RNA, whereas in bone marrow and pancreas (obtained from alternative sources, see “Experimental Procedures”), robust transcripts were amplified under the same PCR conditions (Fig. 3A). In contrast, the Dppa3 source gene primer pair successfully amplified a product of the expected size from NT2 cell RNA, confirming expression of Dppa3 coordinate with the pluripotency-related genes Nanog and Oct4 (Fig. 3B). However, no amplification was observed in bone marrow and pancreas.
Confirmation of RT-PCR specificity for \(Dppa3\) source gene and \(Dppa3(d)\)

<table>
<thead>
<tr>
<th>Primer</th>
<th>Tissue/Cell</th>
<th>Sequenced amplicon(bp)</th>
<th>Identity to (Dppa3) gene %</th>
<th>Identity to (Dppa3(d)) loci %</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Dppa3(d)) (\rightarrow)</td>
<td>Pancreas</td>
<td>334</td>
<td>81</td>
<td>100</td>
</tr>
<tr>
<td>(Dppa3(d)) (\rightarrow)</td>
<td>Pancreas</td>
<td>323</td>
<td>81</td>
<td>100</td>
</tr>
<tr>
<td>(Dppa3(d)) (\rightarrow)</td>
<td>Bone marrow</td>
<td>311</td>
<td>81</td>
<td>100</td>
</tr>
<tr>
<td>(Dppa3(d)) (\rightarrow)</td>
<td>Bone marrow</td>
<td>329</td>
<td>81</td>
<td>100</td>
</tr>
<tr>
<td>(Dppa3) (\rightarrow)</td>
<td>NT2 cells</td>
<td>72</td>
<td>100</td>
<td>ND</td>
</tr>
<tr>
<td>(Dppa3) (\leftarrow)</td>
<td>NT2 cells</td>
<td>79</td>
<td>100</td>
<td>ND</td>
</tr>
</tbody>
</table>

* ND, not determined.

**DISCUSSION**

Germ line transmission is a prerequisite for genes that give rise to processed pseudogenes, and elevated source gene transcription levels further promote such events (21). This primary feature of inherited retrotrotransposition inspired us to focus on pluripotent cell-specific genes as a potential source of processed pseudogenes in the human genome. Previously, we proposed a role of extensive retrotrotransposition as a potential signature event of ES cell-specific genes (4). We found that Nanog, Oct4, and Dppa3 (Stella/PGC7) each possess multiple processed pseudogenes, and our data suggested that retrotrotransposition frequency of ES cell-specific genes far exceeds that of non-ES cell-specific genes. Moreover, we proposed that this “pseudogene signature” may prove useful in identifying novel ES cell-specific genes in future. On this basis, we continued to develop the hypothesis to assess whether any of the ES cell-specific (or pluripotent cell-specific) genes were retrotransposed to form transcriptionally active genomic integrants. First, we identified further examples of processed pseudogenes originating from pluripotent cell-specific genes. The \(Dppa3\) family of genes were identified as a cluster of genes whose developmental expression pattern was similar to that of Oct4 (11). Like Oct4, this cluster of genes also appeared predisposed to the generation of processed pseudogenes, consistent with the original tenet of the study. Next, we progressed to the question of whether any of the putative pluripotency gene-related retro-pseudogenes were transcriptionally active. We hypothesized that the creation of functionally active products of retrotrotransposition that originated from pluripotency-related genes may indicate far reaching implications in evolution.

Here, we reveal that retrotrotransposition of the pluripotent tissue-specific \(Dppa3\) gene created a transcriptionally active daughter genomic sequence, which displays a discrete tissue-specific expression pattern. The functional role (if any) of this descendant transcript remains to be elucidated. As for precedent, the reported functional manifestations of other transcribed retro-pseudogenes include protein expression (22) as well as regulatory mechanisms mediated by duplex mRNA formation with parent transcripts that impact inhibitory (23) or stabilizing effects (24). However, as the tissue distribution of \(Dppa3(d)\) transcript is distinct from its source gene, a role in mRNA duplexing is unlikely. Furthermore, recent studies investigating the multipotency of adult tissue-derived cells have utilized RT-PCR detection of \(Dppa3\), Oct4, and Nanog gene transcripts as markers of pluripotency (25). Given our present finding that a \(Dppa3\)-derived retrogene is expressed in bone marrow, it may be prudent to readdress the use of these pluripotency genes in this context. Moreover, further studies should reveal which specific cell types within bone marrow, peripheral blood, pancreas, thyroid, and adrenal gland express the \(Dppa3(d)\) retrogene, and these ongoing efforts will likely expedite its functional characterization.

Given the increased propensity for highly expressed genes within the germ line to propagate new, functional DNA, a clear advantage is provided for ES cell-specific genes or other pluripotency-related genes to promote genetic diversity and play a key evolutionary role. Whether this is the case for the \(Dppa3\)-derived retrogene, \(Dppa3(d)\), is tempting to speculate, but clarification will require further study.
Acknowledgment—We thank Jee-Hyung Lee for helpful discussions and advice.

REFERENCES

Adult Tissue-specific Expression of a *Dppa3*-derived Retrogene Represents a Postnatal Transcript of Pluripotent Cell Origin
Stephen J. Elliman, Isaac Wu and Daniel M. Kemp


Access the most updated version of this article at doi: 10.1074/jbc.C500415200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 25 references, 5 of which can be accessed free at http://www.jbc.org/content/281/1/16.full.html#ref-list-1