<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Spatial-temporal transcriptional dynamics of long non-coding RNAs in human brain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Zhang, XQ; Wang, Z; Poon, MW; Yang, J</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>Human Molecular Genetics, 2017, v. 26 n. 16, p. 3202-3211</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2017</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/259391">http://hdl.handle.net/10722/259391</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Abstract
The functional architecture of the human brain is greatly determined by the temporal and spatial regulation of the transcription process. However, the spatial and temporal transcriptional landscape of long non-coding RNAs (lncRNAs) during human brain development remains poorly understood. Here, we report the genome-wide lncRNA transcriptional analysis in an extensive series of 1340 post-mortem human brain specimens collected from 16 regions spanning the period from early embryo development to late adulthood. We discovered that lncRNA transcriptome dramatically changed during fetal development, while transited to a surprisingly relatively stable state after birth till the late adulthood. We also discovered that the transcription map of lncRNAs was spatially different, and that this spatial difference was developmentally regulated. Of the 16 brain regions explored (cerebellar cortex, thalamus, striatum, amygdala, hippocampus and 11 neocortex areas), cerebellar cortex showed the most distinct lncRNA expression features from all remaining brain regions throughout the whole developmental period, reflecting its unique developmental and functional features. Furthermore, by characterizing the functional modules and cellular processes of the spatial-temporal dynamic lncRNAs, we found that they were significantly associated with the RNA processing, neuron differentiation and synaptic signal transportation processes. Furthermore, we found that many lncRNAs associated with the neurodegenerative Alzheimer and Parkinson diseases were co-expressed in the fetal development of the human brain, and affected the convergent biological processes. In summary, our study provides a comprehensive map for lncRNA transcription dynamics in human brain development, which might shed light on the understanding of the molecular underpinnings of human brain function and disease.

Introduction
Human brain development is a complex and precisely regulated process, which defines the structural formation and function of the brain. Correlated with this process are highly dynamic transcriptional changes in multiple human brain regions ranging from fetal development through adulthood (1–3). The disruption of this process may not only lead to psychiatric diseases but also increased susceptibility to certain neurological disorders (4,5). Increasing studies have shown that the dysregulation of certain development-associated genes plays important roles in the pathogenesis of the neurodegenerative diseases, including Alzheimer’s disease (AD) (6) and Parkinson’s disease (PD) (7).
Understanding the spatial-temporal transcriptional dynamics underlying human brain development at a systemic level, therefore, is crucial for elucidating the molecular mechanisms involved in the formation, function as well as disorder of the human brain.

Recent studies have begun to characterize the transcriptome of human brain by applying high throughput methods. For example, by profiling the transcriptome of the prefrontal cortex of the human brain using microarray, Colantuoni et al. characterized the temporal brain transcription dynamics in the prefrontal cortex from fetal development through ageing (1). By using a more comprehensive microarray profiling method, Kang et al. explored both spatial and temporal transcriptomic dynamics of the human brain in multiple brain regions and developing stages (2). These studies together with other related investigations (3,8,9) provide important insights into the molecular architecture underlying human brain development, but have a limitation that they either focus mainly on protein-coding genes, or predominantly focus only on very few brain regions or developmental time points.

It has been increasingly accepted that long non-coding RNAs (lncRNAs), a new and large class of regulatory genes without significant protein-coding potential, take the major proportion of human transcriptome and play important regulatory roles in brain development, function and disease (5,10–15). Investigating how genome-wide lncRNA transcription profiles alter spatially and temporally during human brain development, therefore, is important for understanding the roles of lncRNAs in human brain functions and diseases. However, a comprehensive lncRNA transcriptome analysis across human brain development has been lacking, which limit our understanding of the lncRNA roles in human brain greatly.

In the present study, we aimed to comprehensively explore the spatial-temporal transcription dynamics of lncRNAs during human brain development. We explored the lncRNA transcriptome based on whole-genome microarray in a huge cohort of 1340 post-mortem human brain samples including multiple brain regions and across several developmental stages. We investigated the spatial-temporal lncRNA transcription dynamics from the early embryonic development to late adulthood for 16 brain regions (cerebellum cortex, thalamus, striatum, amygdala, hippocampus and 11 neocortex areas). Moreover, we characterized the potential functional modules and networks of the dynamic lncRNAs through exploring their co-expressions with protein-coding genes from the same data set. We also investigated the associations between lncRNA spatial-temporal dynamics with the neurodegenerative Alzheimer’s and Parkinson’s diseases. Our study provides novel insights into the biological relevance of lncRNAs in human brain development, functions and disorders.

Results

Global visualization of lncRNA transcriptome in human brain development

We first characterized the global lncRNA transcription landscape in all samples enrolled. By performing multidimensional scaling (MDS) analysis, a useful and commonly used method for representing high-dimensional data in low-dimensional space but preserving the key similarities between samples, we found that the lncRNA transcription features in the prenatal samples (stages before birth; periods 1–7 in Table 1) were distinct from those in the adulthood ones (adolescence to late adulthood; periods 13–15 in Table 1) in spite of their tissue origins (Fig. 1A).

The postnatal samples (birth to adolescence; periods 8–12 in Table 1), as expected, were distributed widely in the transition area between the prenatal and adulthood samples, but were more adulthood-like in MDS plot (Fig. 1A). In another similar strategy by principal component analysis (PCA), similar findings were observed (Fig. 1B): the prenatal samples were clearly separated from both postnatal and adulthood samples by PCA1, while the latter two largely overlapped. However, when the sample grouping was performed by sex, no significant separation was observed (Supplementary Material, Fig. S1). These findings suggest that temporal factor plays an important role on lncRNA transcription during the human brain development, and that its regulatory effects predominantly happen during the fetal development instead of after birth.

Temporal transcription dynamics of lncRNAs across human brain development

We then investigated how lncRNA transcription was temporally regulated across the whole brain development. For this, we first classified the samples in each of the six main brain regions, which included cerebellum cortex (CBC), mediodorsal nucleus of the thalamus (MD), striatum (STR), amygdala (AMY), hippocampus (HIP) and 11 neocortex (NCX) areas, into three temporal categories: prenatal, postnatal and adulthood (Table 1), and the differentially expressed (DEX) lncRNAs between different temporal groups were then characterized by one-way analysis of variance (ANOVA) in each brain region.

By using the false discovery rate (FDR) < 0.01 and fold change ≥ 2 in ANOVA as the threshold, we identified a total of 309 temporally DEX lncRNAs in CBC, 181 in HIP, 331 in STR, 214 in MD, 248 in AMY and 274 in NCX (Supplementary Material, Table S1). Further analysis revealed that the majority of these temporally DEX lncRNA genes were identified from the prenatal versus postnatal or prenatal versus adulthood comparison, while very few were between the postnatal and adulthood comparison, in spite of the long temporal altitude of the latter two periods (20–82 years old). For example, among the 309 DEX lncRNAs in CBC, only 4 were found between the postnatal and adulthood comparison. Only 2 of 181 DEX lncRNAs in HIP, 5 of 331 in STR, 3 of 214 in MD, 6 of 248 in AMY and 2 of 274 in NCX, were from the postnatal and adulthood comparison, respectively. This is consistent with the above MDS and PCA findings (Fig. 1) that the postnatal and adulthood samples shared many similarities while the prenatal ones were more distinct in lncRNA features. We also observed that among these six sets of temporally DEX lncRNAs, a subset of 37 lncRNAs were commonly DEX among all six main brain regions analyzed (Supplementary Material, Table S2). The un-supervised clustering of all samples by this set of commonly DEX lncRNAs demonstrated the similar classification pattern, in which the prenatal samples were located on the same branch of the dendrogram in spite of their origins, whereas the other major branch was predominantly populated with the samples of both postnatal and adulthood stages (Fig. 2A). Similarly, when the un-supervised clustering was performed in each brain region by the corresponding set of DEX lncRNAs, similar results were observed (Supplementary Material, Fig. S2).

Further inspection of these sets of DEX lncRNAs revealed several specially interesting lncRNA candidates. One of them was lncRNA LINC-PINT, which demonstrated progressive expression increase with age in all six brain regions (Fig. 2B);
LncRNA RP11-675F6.4 demonstrated a similar expression increase pattern in CBC (Fig. 2C), and PAXIP1-AS2 in AMY (Fig. 2D). We also observed the progressive decrease of lncRNA expression with age. For example, lncRNA MLLT4-AS1 showed a progressive decreasing of expression in all brain brains (Fig. 2E).

The age-dependent expression patterns of these lncRNAs across both prenatal and postnatal stages suggest their potential roles in brain developing and ageing.

Table 1. Spatial and temporal distribution of the post-mortem human brain samples enrolled in the study (n = 1340)*

<table>
<thead>
<tr>
<th>Period</th>
<th>Description</th>
<th>Age</th>
<th>HIP</th>
<th>AMY</th>
<th>STR</th>
<th>MD</th>
<th>CBC</th>
<th>Neocortex (NCX)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prenatal 1</td>
<td>Embryonic</td>
<td>4 PCM&lt;Age &lt;8 PCM</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Early fetal-I</td>
<td>8 PCW&lt;Age &lt;10 PCW</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>26</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Early fetal-II</td>
<td>10 PCM&lt;Age &lt;13 PCM</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>Early mid-fetal-I</td>
<td>13 PCM&lt;Age &lt;16 PCM</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>Early mid-fetal-II</td>
<td>16 PCM&lt;Age &lt;19 PCM</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>33</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>Late mid-fetal</td>
<td>19 PCM&lt;Age &lt;24 PCM</td>
<td>13</td>
<td>9</td>
<td>13</td>
<td>13</td>
<td>11</td>
<td>61</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>Late fetal</td>
<td>24 PCM&lt;Age &lt;38 PCM</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>Postnatal 8</td>
<td>Neonatal and early infancy</td>
<td>0 M&lt;Age &lt;6 M</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>Late infancy</td>
<td>6 M&lt;Age &lt;12 M</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>Early childhood</td>
<td>1 Y&lt;Age &lt;6 Y</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>31</td>
<td>14</td>
</tr>
<tr>
<td>11</td>
<td>Middle and late childhood</td>
<td>6 Y&lt;Age &lt;12 Y</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>Adolescence</td>
<td>12 Y&lt;Age &lt;20 Y</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>Adulthood 13</td>
<td>Young adulthood</td>
<td>20 Y&lt;Age &lt;40 Y</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>72</td>
<td>27</td>
</tr>
<tr>
<td>14</td>
<td>Middle adulthood</td>
<td>40 Y&lt;Age &lt;60 Y</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>29</td>
<td>11</td>
</tr>
<tr>
<td>15</td>
<td>Late adulthood</td>
<td>60 Y&lt;Age</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>28</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>82</td>
<td>76</td>
<td>85</td>
<td>80</td>
<td>77</td>
<td>442</td>
<td>163</td>
</tr>
</tbody>
</table>

PCW, post-conceptual weeks; M, postnatal months; Y, postnatal years; HIP, hippocampus; AMY, amygdala; STR, striatum; MD, mediiodorsal nucleus of the thalamus; CBC, cerebellar cortex; PFC, prefrontal cortex; PC, parietal cortex; TC, temporal cortex; OC, occipital cortex. *The detailed description of the clinical characteristics of the samples can be found at (2).

Figure 1. Global view of the lncRNA transcription pattern in human brain. (A) Multidimensional scaling (MDS) analysis; (B) Principal component analysis (PCA). In both MDS and PCA plotting, prenatal brain samples were clearly separated from both postnatal and adulthood ones.

Spatial difference of lncRNA transcription in human brain

We further explored the spatial difference of lncRNA transcription between various brain regions. To do so, the lncRNA expression profiles between the six main brain regions were compared by one-way ANOVA within each developmental stage (prenatal, postnatal and adulthood), separately. Supplementary Material, Table S3 lists the DEX lncRNAs in each comparison analysis.
Figure 2. Temporal dynamics of lncRNA transcriptome across human brain development. (A) Un-supervised clustering analysis of all samples based on the commonly DEX lncRNAs revealed the similarities of postnatal and adulthood samples on lncRNA transcription features as compared with the prenatal ones. (B–E) Age-dependent expression patterns of representative lncRNAs.
We observed that the prenatal samples demonstrated the fewest spatial lncRNA differences, while the postnatal and adulthood samples had increasingly more region-biased lncRNA expressions. Of the 7767 expressed lncRNAs analyzed, only 249 lncRNAs were found to be regionally differentially expressed (FDR < 0.01 and fold change ≥ 2) during the prenatal stage. On contrast, there were 469 and 496 DEX lncRNAs identified in the postnatal and adulthood stage, respectively. Of particular note, among the two sets of spatially DEX lncRNAs in postnatal and adulthood samples, 358 lncRNAs overlapped (highlighted in Supplementary Material, Table S3), suggesting again the similarity of lncRNA expression features between the postnatal and adulthood brain samples. We also found that CBC showed the most distinct lncRNA transcriptome feature among the six brain regions during the whole developmental period. The abundant expressions of lncRNAs were always observed in CBC but not in others or vice versa. For example, RP11-491F9.1 and LINC01036 demonstrated ultra-high expressions in CBC, but low expressed in the remaining five regions (Fig. 3A and B). These findings suggest that lncRNAs have potentially important roles in deciding the structure or function specificity of the human brain.

To test the generality and robustness of our analysis, we then validated the spatially DEX lncRNAs in an independent
data set (features detailed in Supplementary Material, Table S3), which were profiled on the same microarray platform with the data cohort analyzed above. The 358 commonly DEX lncRNAs between the postnatal and adulthood stages were chosen for confirmation, considering the sample availability of the testing data (without prenatal samples). As shown in Figure 3C and D, the un-supervised clustering of both training- (Fig. 3A) and testing-set (Fig. 3D) samples by the 358 spatially DEX lncRNAs from the training data set could clearly separate the samples into several distinct groups based on their tissue origins. In this, CBC samples were again clearly separated from other samples and demonstrated the most distinct lncRNA expression pattern, indicating that the regional DEX lncRNAs identified here is representative and may reflect the anatomical differences of different human brain regions.

Function prediction of lncRNAs through weighted gene co-expression network analysis (WGCNA)

To determine the potential functional categories and cellular processes of lncRNAs in the human brain, we performed WGCNA on lncRNAs and mRNAs which were derived from the same microarray data set, followed by annotating the co-expressed mRNAs using gene ontology (GO) enrichment analyses. With default parameters (inflation value = 1.6), a total of 23 lncRNA-mRNA co-expression modules were identified, and 22 of which were with at least three lncRNA genes (module 23 excluded; Fig. 4A and B). All of these modules were significantly enriched for at least one GO BP term ($P < 0.05$, Supplementary Material, Table S4). We named each module after their most significant enriched function.

The biggest module was RNA-processing module (module 1, M1). M1 contained 2629 mRNAs and 1342 lncRNAs. Among the 1342 lncRNAs, 116 were found to be temporally DEX lncRNAs and 94 were spatially DEX lncRNAs identified above. GO enrichment analysis showed that M1 was mainly enriched in the biological processes of RNA processing ($P = 1.49E-27$), cellular compound metabolism ($P = 1.58E-23$), gene expression ($P = 1.63E-22$), and ribonucleoprotein complex biogenesis ($P = 1.18E-21$) (Fig. 4C). Another associated large module was module 3 (M3). This module contained 1663 mRNAs and 375 lncRNAs. Among these 375 lncRNAs, 164 (44%) were found to be temporally DEX lncRNAs. GO enrichment analysis of this module showed that M3 was enriched for synaptic signal transportation processes, including synaptic transmission ($P = 3.85E-10$), ion transport ($P = 2.91E-11$), chemical homeostasis ($P = 3.85E-10$) and cell-cell signaling ($P = 4.04E-09$) (Fig. 4C). In the eigengenes heatmap, M1 tended to have high abundant expression in prenatal samples as relative to postnatal and adulthood ones, but M3 showed opposite expression pattern (Fig. 4B).

We also observed an interesting module, M14, which was associated with the spatial-based lncRNA difference. This module was comprised of a total of 108 mRNAs and 31 lncRNAs. Of which, 29 lncRNAs (93.5%) were found to be spatially DEX lncRNAs. Furthermore, these lncRNAs tended to be highly expressed in the CBC area while low expressed in other brain areas (Fig. 4B), which included RP11-491F9.1 and LINC01036 identified above (Fig. 3A and B). In the co-expression network, they were closely associated with the key neuronal development-associated mRNAs, e.g. CNPY1 and EN2 (Fig. 4D). GO analysis revealed that M14 was enriched with neural differentiation (Fig. 4C). For example, it was involved in the retina layer formation ($P = 1.7E-05$), spinal cord associated neuron differentiation ($P = 4.4E-05$), cell differentiation in hindbrain ($P = 7.2E-05$) and cerebellar cortex formation ($P = 7.2E-05$).

Mapping spatial-temporal dynamic lncRNAs to AD and PD candidate genes

Finally, we investigated the associations between the spatial-temporal lncRNA transcription dynamics and the development of neurodegenerative disorders AD and PD. We mapped the spatial-temporal dynamic lncRNAs to the disease candidate lncRNAs of AD and PD, respectively, which were derived from the published microarray studies (details described in Materials & Methods section).

In AD, among the 207 disease candidate lncRNAs (absolute fold change between AD and normal brain controls $\geq 2$; Supplementary Material, Table S5), 51 (24.6%) were found to overlap with the temporally DEX lncRNAs, and 29 (14%) were overlapped with the spatially DEX ones (Supplementary Material, Table S5). By mapping the overlapping lncRNAs to the WGCNA modules identified above, we found that they were significantly enriched in M1, M3, M4, M6 and M9 modules. Amongst, M1 and M4 tended to have increased gene expression at the prenatal stage than that after birth, while M3, M6 and M9 demonstrated opposite temporal change pattern (Fig. 4B). In the GO analysis, they were grouped into two main categories: RNA metabolism which included M1 and M6 (Fig. 4C), as well as neuronal signal transportation which included M3, M4 and M9 (Fig. 5A).

In PD, among the 56 dysregulated lncRNAs between PD brains and normal controls (absolute fold change between PD and normal brain controls $\geq 1.5$; Supplementary Material, Table S5), 20 of them (35.7%) overlapped with the temporally DEX lncRNAs (Supplementary Material, Table S5), and 6 (3%) with spatially DEX lncRNAs (Supplementary Material, Table S5). In WGCNA, they were enriched in M1, M4 and M9 modules, which contained 9, 5 and 6 overlapping lncRNAs, respectively, and were also involved in the GO categories of RNA metabolism and neuronal signal transportation similarly (Fig. 5A).

Discussion

Here, we reported the investigations of genome-wide lncRNA transcriptional profiles in large cohort of post-mortem human brain tissues including both the prenatal and postnatal developmental stages and for multiple brain regions. We found that lncRNA transcriptome experienced dramatic spatial-temporal dynamics during the human brain development, and were associated with the RNA processing, neuron differentiation and synaptic signal transportation processes. We also found that there were significant associations between the lncRNA spatial-temporal transcription dynamics and the development of Alzheimer’s and Parkinson’s diseases, and many lncRNAs associated with the neurodegenerative disorders were also co-expressed during the fetal development. Our study provides an important step for understanding the molecular and functional mechanisms of lncRNAs in brain development, function and neurodegenerative disease.

As a novel and large class of regulatory genes, lncRNAs have been increasingly indicated in brain function and disease (5,10–17). They have also been implicated in neuronal development and even intellectual disability (18). Here, we observed that the most striking temporal changes of lncRNA transcriptome occurred in the fetal development, and surprisingly little
difference was observed after birth till the late adulthood despite their large temporal interval. The hypothesis is that the developing fetal brain would have more genetic or epigenetic differences, both due to the possibility of stem cell maintenance and differentiation or because of the neuronal connectivity establishment that occurs during the fetal period. A similar neonatal gene expression shift in brain development has been previously observed for mRNAs in both human and primate (9,19).

We also observed that IncRNA transcriptome was different among samples from different brain regions, and that this region-biased expression becomes increasingly more obvious as the increasing ageing towards the late adulthood. In this, fewer IncRNA expression differences were observed during prenatal...
development than that in postnatal and adult stages. In spite of this, CBC demonstrates the most distinct lncRNA expression profile from other five regions (NCX, AMY, HIP, STR and MD) across the whole developmental stage. This may reflect the dramatic developmental, cellular, and functional differences between CBC and forebrain structures.

Several lncRNAs that demonstrated special spatial-temporal expression patterns have previously been implicated in the brain-associated cellular processes. One of such instances is LINC-PINT, which demonstrated age-dependent expression increase in all six brain regions here, has been characterized as a p53-induced lncRNA in previous studies (that is how it gets its name) (20). It can interact with PRC2 and regulate the expression of genes involved in the TGF-β, MAPK and p53 pathways, and is implicated in senescence, ageing, and age-related diseases (20–22). For the lncRNA RP11-675F6.4 that is increasingly expressed with age in CBC, and PAXIP1-AS2 in AMY, as well as the CBC-specific lncRNA RP11-491F9.1 and LINC01036, their functions in brain have not been reported yet. Their specific expression patterns identified here warrant the further study.

To determine how the spatial-temporal DEX lncRNAs function, we performed weighted gene co-expression network analysis (WGCNA) between the co-expressed lncRNAs and mRNAs followed by GO functional enrichment analysis. We found that the temporally DEX genes were largely associated with the process of RNA processing or metabolism (e.g. the biggest M1 and M6), which is in agreement with the previous findings that lncRNAs play important roles in RNA processing and splicing, which helps the brain produce extraordinary gene transcription diversity, and is also implicated in disease (23). At the same time, as expected, they were also significantly associated with the neuron-specific biological processes, including synaptic signal transportation (e.g. M3) or other types of signal perception of neurons (e.g. M9). On contrast, the spatially DEX genes were highly enriched for neuronal cell differentiation categories. For example, M14, which was enriched with lncRNAs that were overexpressed in CBC than in other brain regions, was found to be mainly associated with the neuronal cell differentiation, which suggests that cell differentiation and specification are potentially important ways to decide the function diversity of different brain regions.

By mapping the spatial-temporal dynamic lncRNAs to the neurodegenerative disorders AD and PD, it was interesting to find that many lncRNA genes associated with AD and PD, actually were co-expressed during the human fetal development, affecting the convergent molecular pathways. There were 38.6% and 38.7% disease lncRNAs in AD and PD, respectively, were the spatial-temporal DEX lncRNAs during the human brain development at the same time. Moreover, they were distributed in the key lncRNA-associated modules that were important for the brain development, which included M1, M4, M6 and M9, suggesting that the human brain fetal development and neurodegenerative diseases may happen through the similar cellular processes and signal pathways.

While this study may provide useful implications for the future basic and clinical study on human brain function and disorders, it has several limitations. First, the transcription analysis of the current study was performed on the post-mortem tissues containing mixed cell types, which therefore cannot provide information for individual specific cell type or layer. Second, we used conservative threshold to identify DEX lncRNAs through the whole study, and this may lead to the possibility that some important properties of the smaller sample population may be overlooked. Finally, the experimental validation and functional exploration of specific lncRNAs with special spatial-temporal expression patterns are lacking. Therefore, a much deeper study at single-cell level and with experimental function exploration in the future is still needed. Nevertheless, apart from the limitations of this study, we do set an important step towards the understanding of the temporal-spatial transcriptional profiling of lncRNAs on human brain development and disorders, we hope that our study would shed light on the future studies on the corresponding mechanistic studies.

Materials and Methods

Post-mortem human brain transcriptome data by Affymetrix exon 1.0 ST array

The transcription data of a large cohort of anatomical comprehensive post-mortem human brain samples on Affymetrix Exon 1.0 ST array were obtained from the gene expression omnibus (GEO). This huge data cohort contains a total of 1340 post-mortem human brain samples collected from 57 developing and adult donors without remarkable clinical symptoms or large-scale genomic abnormalities, and spans the periods from early embryogenesis to late adulthood (2). For each of the donor, the specimens were collected from up to 16 different anatomical brain regions, including the cerebellar cortex (CBC), mediodorsal nucleus of the thalamus (MD), striatum (STR), amygdala (AMY), hippocampus (HIP) and 11 neocortex (NCX) areas. The detailed spatial and temporal distribution of the samples is described in Table 1.

Affymetrix Exon 1.0 ST array data processing and lncRNA retrieving

The raw microarray data (CEL file) on Affymetrix Exon 1.0 ST platform, which include 1.4 million probe sets and have a comprehensive coverage of the human genome, were pre-processed and log2 transformed using the Affymetrix Power tools (APT) with command (apt-probeset-summarize -a rma-sketch). The custom chip description file (CDF) for GENCODE genes were downloaded from the BRAINARRAY website (version 20.0.0). The resulted expression values were further filtered by...
calculation of the ‘detection above background metric’ using APT. Only the genes with robust expression (log2-transformed signal intensity ≥ 6 in over 15% of samples), were defined as ‘expressed genes’ and remained for following analysis.

At the same time, we mapped the gene-level probe sets to human GENCODE annotation (version 25) using the custom Perl scripts. Only the probe sets which were annotated as ‘long non-coding RNAs’ and ‘protein-coding genes’ in GENCODE databases were retained, and all others were filtered out. Finally, the expression values for a total of 7677 ‘expressed lncRNAs’ and 13305 ‘expressed protein-coding RNAs’ were obtained for further analysis.

Multiple dimension scaling (MDS) and principal component analysis (PCA)

MDS and PCA analyses were performed by using the isoMDS and prcomp functions from MASS package in R, respectively, to make a global view of lncRNA transcription on all samples. Both methods can reduce the complexity of the high-dimensional microarray data and intuitively visualize the global patterns of objects in a low-dimensional space but use different algorithms (24,25). In both MDS and PCA plots, objects that are clustered in a neighborhood indicate that they have similar properties (24,25).

Differentially expressed (DEX) lncRNA identification

The differentially expressed (DEX) lncRNA genes between different developmental stages or spatial regions were analyzed by one-way analysis of variance (ANOVA) in R. P values from ANOVA were then corrected for multiple comparisons using the Benjamini and Hochberg false discovery rate (FDR) method (26). In ANOVA, genes with a FDR < 0.01 and fold change ≥ 2 were defined as the statistically significant DEX genes.

Weighted gene co-expression network analysis (WGCNA)

We performed WGCNA on lncRNAs as well as mRNAs, which were derived from the same data set, to predict the functional entities of lncRNAs by using the WGCNA R package (27). WGCNA is a systems biology method for describing the correlation patterns among genes across the microarray samples (Langfelder and Horvath, 2008). In WGCNA, genes with similar expressions will be clustered together based on their expression similarities in an unbiased manner, and then be summarized as a module (27). In our study, to remove redundancy, modules with highly correlated module eigengenes (correlation >0.85) were merged together. The top 50 genes expressing the highest within module connectivity were selected and exported to VisANT for network visualization (28). We used Gene Ontology (GO) enrichment analysis implemented in WGCNA package for GO identification.

Mapping DEX lncRNAs to Alzheimer and Parkinson disease candidate genes

To identify the associations between the spatial-temporal dynamic lncRNAs and the neurodegenerative disorders, Alzheimer’s disease (AD) and Parkinson’s disease (PD), we mapped the spatially-temporally DEX lncRNAs identified here to the AD and PD candidate lncRNAs which were derived from the published microarray studies (29,30). The microarray data processing, lncRNA retrieving and expression profile comparing between the normal and disease status in these two studies were performed by using our previously reported method (31). The disease candidate lncRNAs in AD were defined if their expression changes between disease status and normal control were larger than 2 folds, while in PD the cutoff was relaxed to 1.5 folds because of the smaller patient number of the data cohort obtained.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

We thank the authors who generously shared their data.

Conflict of Interest statement. None declared.

Funding

Seeding Project Fund at School of Medicine, South China University of Technology (No. xyy2016005); The Ministry of Science and Technology of China (No. 2017YFA0504400); The National Natural Science Foundation of China (No. 31370791 and 91440110); Funds from Guangdong Province (No. S201201001510, S2013010012457); The project of Science and Technology New Star in Zhubiang Guangzhou city (No. 2012z2200025); Fundamental Research Funds for the Central Universities (No. 2011330003161070, 14lgjc18, 2017MS071).

References


