A Cross-Study Transcriptional Analysis of Parkinson’s Disease

Greg T. Sutherland1,*, Nicholas A. Matigian1,*, Alistair M. Chalk1, Matthew J. Anderson1, Peter A. Silburn1,2, Alan Mackay-Sim1, Christine A. Wells1, George D. Mellick1,3,4,*

1 National Centre for Adult Stem Cell Research, Eskitis Institute for Cell and Molecular Therapies, Griffith University, Brisbane, Queensland, Australia, 2 The School of Medicine, University of Queensland, Brisbane, Queensland, Australia, 3 Department of Neurology, Princess Alexandra Hospital, Brisbane, Queensland, Australia, 4 Department of Neurology, Royal Brisbane and Women’s Hospital, Brisbane, Queensland, Australia

Abstract

The study of Parkinson’s disease (PD), like other complex neurodegenerative disorders, is limited by access to brain tissue from patients with a confirmed diagnosis. Alternatively the study of peripheral tissues may offer some insight into the molecular basis of disease susceptibility and progression, but this approach still relies on brain tissue to benchmark relevant molecular changes against. Several studies have reported whole-genome expression profiling in post-mortem brain but reported concordance between these analyses is lacking. Here we apply a standardised pathway analysis to seven independent case-control studies, and demonstrate increased concordance between data sets. Moreover data convergence increased when the analysis was limited to the five substantia nigra (SN) data sets; this highlighted the down regulation of dopamine receptor signaling and insulin-like growth factor 1 (IGF1) signaling pathways. We also show that case-control comparisons of affected post mortem brain tissue are more likely to reflect terminal cytoarchitectural differences rather than primary pathogenic mechanisms. The implementation of a correction factor for dopaminergic neuronal loss predictably resulted in the loss of significance of the dopamine signaling pathway while axon guidance pathways increased in significance. Interestingly the IGF1 signaling pathway was also over-represented when data from non-SN areas, unaffected or only terminally affected in PD, were considered. Our findings suggest that there is greater concordance in PD whole-genome expression profiling when standardised pathway membership rather than ranked gene list is used for comparison.

Introduction

Parkinson’s disease (PD, OMIM: #168600) is a uniquely human disease that is clinically characterised by cardinal motor symptoms such as postural instability, bradykinesia and resting tremor [1]. In the PD brain there is pathognomonic loss of more than 50% of the dopaminergic neurons in the substantia nigra pars compacta (SNc) located in the midbrain [2,3]. However, the disease is also characterised by non-motor symptoms such as sleep disorders, depression, hyposmia and autonomic dysfunction [4–7]. Accordingly, pathology in PD is not just restricted to the SNc but also affects the olfactory pathway, spinal cord and dorsal cranial nuclei of the medulla [8–11].

High throughput discovery platforms, such as microarrays, that assume no a priori aetiological hypotheses, promise much in elucidating the pathogenesis of complex diseases such as PD. Moreover, one would hope that these microarray data would reveal clues previously inaccessible via other means. Several microarray-based studies have used human tissue to look for differentially expressed genes in Parkinson’s disease [12–21]. The majority of these used post-mortem whole brain tissue from the substantia nigra (SN) [14–17] and although most authors emphasised differential expression in the ubiquitin-proteasome system or cellular energy pathways, their published gene lists appeared quite discordant. Others extended their studies to include pathologically normal brain regions [12,18,20,21] and these highlighted other biological mechanisms such as G-Protein-coupled receptor signaling and transcriptional regulation.

One study assayed SNc dopaminergic neurons only, following isolation by laser capture microscopy (LCM) [13]. Here gender differences were more pronounced than PD versus control differences. Uniquely, one study compared the transcriptomes of whole blood samples [19] and reported expression differences in a number of unrelated genes.

Given an apparent lack of concordance in published data sets one might ask what relevance these transcriptional approaches can have to PD pathogenesis. Certainly the utilisation of post mortem brain tissue appears to represent the best opportunity for finding PD-specific changes in gene expression. Furthermore such ‘benchmarks’ facilitate the evaluation of clinical samples and model systems for their utility in PD research. However the approaches to generating and analysing microarray data are not standardised and therefore could account for much of the apparent discrepancy between reported gene lists.
Pathway analysis of SN data sets reveals common dysregulation of dopamine signaling

Our common analysis method was applied to the 13 datasets from seven analyses that that our platform inclusion criteria, and provided 13 case-control datasets comparing primary tissues derived from PD patients and controls (Table 1). Of these, seven were included because they used the Agilent chip technology. We were also unable to source raw data from the study of Papapetropoulos and colleagues [18], who used Agilent technology. We were also unable to source raw data from the study of Papapetropoulos and colleagues [18].

Table 1. Summary of tissue used, size and array feature of published Parkinson’s disease microarray publications.

<table>
<thead>
<tr>
<th>First Author</th>
<th>PMID</th>
<th>Database Accession</th>
<th>Tissue used in study</th>
<th># of PD cases</th>
<th># of controls</th>
<th>NPDC</th>
<th>Array Type</th>
<th>Summarisation &amp; normalisation Methods</th>
<th>Statistical test used to determine differential expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grunblatt</td>
<td>15455214</td>
<td>-</td>
<td>SN,CB</td>
<td>7</td>
<td>7</td>
<td>-</td>
<td>HG-Focus</td>
<td>MASS</td>
<td>Wilcoxon rank-sum test; fold-change cut-off</td>
</tr>
<tr>
<td>Hauser</td>
<td>15956162</td>
<td>-</td>
<td>SN</td>
<td>6</td>
<td>5</td>
<td>3 (2 PSP, 1 FTDP)</td>
<td>U133A</td>
<td>MASS</td>
<td>2-sample t test</td>
</tr>
<tr>
<td>Vogt</td>
<td>16626704</td>
<td>-</td>
<td>OCTX, CB, PT</td>
<td>4</td>
<td>4</td>
<td>4 MSA</td>
<td>U133A</td>
<td>RMA</td>
<td>ANOVA</td>
</tr>
<tr>
<td>Zhang</td>
<td>15965975</td>
<td>-</td>
<td>SN, PT, BA9</td>
<td>15</td>
<td>19</td>
<td>-</td>
<td>U133A</td>
<td>RMA</td>
<td>ANOVA+ FDR</td>
</tr>
<tr>
<td>Miller</td>
<td>16143538</td>
<td>-</td>
<td>SN, STR</td>
<td>68 &amp; 64</td>
<td>88 &amp; 84</td>
<td>-</td>
<td>CGElink</td>
<td>CGE-RMA+MAS+PLIER</td>
<td>two-class unpaired+FDR</td>
</tr>
<tr>
<td>Moain</td>
<td>16349456</td>
<td>GSE8397</td>
<td>LSN, MSN, SFG</td>
<td>15 &amp; 89 &amp; 82</td>
<td>88 &amp; 76</td>
<td>-</td>
<td>U133A+R</td>
<td>GC-RMA+MAS+PLIER</td>
<td>two-class unpaired+FDR</td>
</tr>
<tr>
<td>Lesnick</td>
<td>17571925</td>
<td>GSE6261</td>
<td>SN</td>
<td>16</td>
<td>9</td>
<td>-</td>
<td>U133 plus 2.0</td>
<td>MASS</td>
<td>ANOVA</td>
</tr>
<tr>
<td>Scherzer</td>
<td>17215369</td>
<td>GSE5613</td>
<td>whole blood</td>
<td>50</td>
<td>22</td>
<td>AD, PSP, MSA, CBD, ET</td>
<td>U133A</td>
<td>MASS</td>
<td>LOOCV</td>
</tr>
<tr>
<td>Castelvetri</td>
<td>17412603</td>
<td>E-MEXP-1416</td>
<td>LCM DA-SN</td>
<td>8</td>
<td>8</td>
<td>-</td>
<td>X3P</td>
<td>GC-RMA</td>
<td>SAM</td>
</tr>
<tr>
<td>Bossers</td>
<td>18462474</td>
<td>-</td>
<td>SN, PT, CN</td>
<td>3</td>
<td>4</td>
<td>1 PD/DEM</td>
<td>Agilent 2K</td>
<td>Loess</td>
<td>Linear regression with FDR</td>
</tr>
</tbody>
</table>

be dissected out of complex transcriptomic data but that these changes are robustly reproducible between comparable studies (Table 4 and Table S2 A–E for full ranked pathway lists).

Neuronal loss in PD is more severe in the lateral SN compared to the medial SN [23–25] while areas such as the superior frontal gyrus (SFG) are largely unaffected. A comparison of these three anatomical areas from the data in Moran et al. (LSN v MSN v SFG) showed a direct correlation between fold changes in genes of dopamine receptor signaling pathway and the severity of PD neuronal loss [14]. Therefore, the prominence of this pathway appeared to represent the disparate numbers of residual dopaminergic neurons in PD compared to control brains.

Correcting for dopaminergic neuronal loss changes ranked pathways

In order to bias our analysis towards underlying pathogenic mechanisms rather than terminal pathology, we devised a correction paradigm based on Moran’s observations on neuronal loss. Our rationale is described in detail in Table S4. 217 genes with fold changes LSN>MSN>SFG from the Moran et al. data sets were defined as potentially “neuronal-loss-associated”, as these were the genes whose differential expression most likely resulted from relative loss of dopaminergic neurons from the brain regions sampled, rather than transcriptomic differences in residual cells. However eight of these genes could be shown to be differentially expressed in residual dopaminergic neurons (Cantuti-Castelvetri et al. study)[13] and so were retained in the analysis. The removal of 209 genes (Table S5) from the SN ranked lists resulted in two alternative pathways gaining prominence: ephrin receptor signaling (p-values, 0.003–0.04) and the axonal guidance pathway (p-values, 0.004–0.049) (Table 4 and Table S6 A–E for full ranked pathway lists).

Assessment of glial contribution to the neuronal-loss corrected datasets

Given the loss of dopaminergic cells in the PD SN the major contribution to the expression profile in the SN PD samples would presumably now come from the non-dopaminergic cells. Furthermore neuropathology in the PD SN is characterised by a reactive gliosis or “glial inflammation” (reviewed by Orr et al., 2002) [26]. Therefore glial markers and in particular reactive microglia markers such as CD68 and ICAM-1 might be expected to be upregulated in the SN data sets [26,27]. However these genes and those of glial markers in general were largely indifferent between PD and controls. This potential anomaly is illustrated further using a selection of PD-related glial markers (Table S7).

Analysis of non-SN brain tissue highlights growth factor signaling pathways

We also analysed non-SN tissues as they are not subject to cytoarchitectural changes seen in the SN or are only affected late in the disease. Five IPA pathways were overrepresented in three out of...
Discussion

Microarrays promise much in elucidating the pathogenesis of complex diseases such as PD but the lack of concordance in published data sets to date certainly questions their relevance. Here we have shown that a standardised approach to analysing PD-related microarray data can account for a considerable proportion of the discordance. We used a common analytical approach which improved data convergence and uncovered new leads for PD pathogenesis. We also recognised a potential anatomical bias in the datasets derived from brain regions with high neuronal loss. Our approach therefore provided an improved comparative analysis between existing datasets and further considered ‘tissue-of-origin’ effects.

Pathway analysis may uncover concordance between datasets not found in gene lists

Complex phenotypes, by their very nature, are aetiologically heterogeneous. This implies that single gene signatures may not be shared by all affected individuals. However, the identification of particularly relevant genetic pathways, have a higher probability of being revealed as convergent across multiple individuals and multiple studies than individual genes per se. Moreover, the differences reflected in a pathway or network of genes may be robust enough to overcome the effects of experimental noise and inter-study variability prone to bias single gene expression values. Therefore, this approach results in an increase in sensitivity to detect interesting and novel patterns in gene expression between multiple samples of cases and controls.

Common data analysis highlights dopamine signaling pathway in SN

Context is very important in gene expression studies and as expected the analysis of SN tissue-derived data sets further improved our concordance and highlighted the ‘dopamine receptor signaling’ pathway. However rather than representing a primary pathogenic effect the extensive down regulation of genes such as DOPA decarboxylase (DDC), dopamine receptor 2 (DRD2), dopamine transporter (SLC6A3 or DAT) and tyrosine hydroxylase (TH) was probably entirely due to a disproportionate number of SN dopaminergic neurons between cases and controls. On a positive note, the microarray data was providing accurate molecular fingerprints of the comparative tissue being examined; a
modern pixelated analogy to the histological section. However the relative neuronal loss in the SN also reduced our signal to noise ratio for pathogenic relevance.

Accordingly we pursued two alternative approaches to maximise potentially useful information on the underlying biological processes. First, the implementation of a correction factor for dopaminergic neuronal loss in the SN data sets and second, the analysis of non-SN or unaffected tissues for data convergence. The purpose of the correction factor was not to magically recreate the early disease landscape but to remove ‘red herrings’ that solely reflected the relative numbers of neurons between cases and controls. The retention of genes differentially regulated in the residual dopaminergic neuron data set should have improved the overall specificity of this approach.

Following our ‘neuronal loss’ correction, two alternative pathways gained prominence: ephrin receptor signaling and the axonal guidance pathway. The latter is consistent with the findings of recent studies [12,28] including one that combined gene expression data with genotype data from two genome-wide association studies [28]. There is actually considerable overlap between the axon guidance and ephrin signaling pathways with ephrins along with netrins, slits and semaphorins being the main families of guidance molecules in the developing nervous system [29]. It remains to be clarified whether the differential expression of axon guidance genes in PD represent neurodevelopmental manifestations, compensatory attempts at rewiring, or dysregulated expression patterns induced by a devastated environment.

As discussed above microarray data is very powerful in illustrating cytoarchitectural differences between cases and controls such as dopaminergic neuron loss. Given the considerable literature supporting the involvement of glia in PD pathogenesis [26,27,30,31] we might have expected glial markers to inversely correlate such as dopaminergic neuron loss. Given the considerable literature supporting the involvement of glia in PD pathogenesis [26,27,30,31] we might have expected glial markers to inversely correlate with the SN data sets. The absence of significant fold changes in activated microglial markers in particular argues against the underlying biological difference between cases and controls but to understand some of the early pathogenic processes, we would ideally want to assay a brain region very similar to SN but that is only belatedly affected.

An additional consideration is the ability of the pathway approach, used in our analyses, to provide adequate specificity for PD over other neurodegenerative conditions. This issue remains to be clarified, and requires further investigation. It is important to recognise that there may be genetic expression patterns common to neurodegenerative diseases, generally. These may reflect common pathological changes (such as cell death, markers of oxidative stress or neuro-inflammation etc) or shared risk factors influencing neurodegeneration. There are still inherent difficulties in obtaining reproducible gene expression data from post mortem brain, even if an optimal region of the brain could be assayed [41–44]. Furthermore this information can only be used retrospectively for the potential benefit of future PD patients. Therefore there is considerable interest in developing strategies to obtain human RNA from more accessible sources such as blood or neuronal-like cell lines. The finding of down regulation of β-synuclein in microarray analysis of whole blood samples from PD patients versus controls is exciting because it implies that PD-specific changes can be found ante mortem in readily accessible tissues [19]. It is also of note that these tissues also revealed differential expression of the VEGF signaling pathway (observed in other non-SN tissue samples). However, there is a real risk that peripheral tissues, such as whole blood examined here, may express few proteins fundamental to the disease process and therefore be of limited ability to demonstrate case - control differences relevant to nervous system disorders [45]. The lack of available gene expression data from multiple tissues in PD patients at various stages of the disease prevents such an analysis but highlights the need for ongoing research efforts in this area.

Interestingly one peripherally accessible neural tissue, the olfactory mucosa, has been used to demonstrate significant differences in functional assays and gene expression between schizophrenics, bipolar affective disorder and controls [46]. Such cells from PD patients and controls may yet provide an opportunity to interrogate neuronal mechanisms without relying on post mortem tissue [47].

In this paper we have presented a summary of the available microarray data from PD case-control studies and have suggested some potential strategies for uncovering primary pathogenic mechanisms. For others who wish to use and explore these data we have constructed an online database which enables rapid evaluation on a single gene or pathway basis.

**Materials and Methods**

We conducted literature searches in National Center for Biotechnology Information (NCBI) PubMed and dataset searches...
in NCBI gene expression omnibus and ArrayExpress (EBI) [48,49] to identify all reported microarray studies that explored differential gene expression in Parkinson disease. Studies satisfied the inclusion criteria if they: 1) compared tissues from PD patients and controls; 2) assessed transcripts on a genome-wide basis; and 3) used Affymetrix gene expression arrays. For the selected studies we obtained raw microarray data (CEL files) from public microarrays repositories [13,14,19] or from the study authors [16; Vogt, 2006 #1529,21,]. In one case the data was publicly available from the follow up study [28] rather than the primary study [18].

**Meta-analysis data summarisation, normalisation and analysis of variance**

All studies used Affymetrix arrays, the probes on the arrays and the experimentally chosen fluorescence thresholds varied. Consequently, the data could not be simply combined without avoiding study bias and the effects of probe-level sequence information [50]. To overcome this problem, the raw data (as CEL files) for each dataset were imported individually into GeneSpring™ 7.3.1(Agilent) and the probes sets were summarised and normalised by the Robust Multichip Average (RMA) algorithm[51]. Some studies included non-PD disease controls but our analysis was performed on PD and control patients only (Table 1). For studies which used multiple brain regions, each area was treated as a separate data set. Differentially expressed genes between PD and controls were determined by an analysis of variances (ANOVA) using a Welch t-test with a p-value cut-off of ≤0.01.

**Pathway over-representation analysis**

The ranked genes lists for each study were assessed by integrating the data at a pathway level. Each ranked list was imported into the Ingenuity Pathways Analysis 6.3 (IPA, from Ingenuity® Systems, www.ingenuity.com) which incorporates an extensive literature-derived knowledge base from which to assign pathway affiliation. The significance value for pathway over-representation was calculated using a right-tailed Fisher’s exact test. Each pathway was ranked by assessing the number of studies that were statistically over-represented (p-value ≤0.05). This pathway over-representation ranking was performed individually on data sets utilising substantia nigra (SN) [Hauser, Moran LSN+MSN, Zhang, Lesnick][14,16,21,28] and non-SN tissues (Zhang-Putamen, Zhang-Ba9, Moran-SFG, Vogt-OCT, Vogt-Putamen, Vogt-CB)[14,20,21]. Additionally this analysis was performed independently on data sets derived from whole blood (Scherzer)[19] and dopaminergic neurons (Cantuti-Castelvetri)[13].

**Correction for dopaminergic neuronal loss**

The substantia nigra pars compacta of PD patients is characterised by the loss of neuromelanin-containing dopaminergic neurons [2,3]. Furthermore neurons are lost in a particular pattern; severity decreasing from ventrolateral to dorsomedial [23–25]. In an attempt to correct for the effects on expression arising directly from the neuronal loss associated with PD, we devised the following correction paradigm. We first used data from the Moran et al. study[14] which compared gene expression profiles in three brain regions, lateral SN (LSN), medial SN (MSN) and superior frontal gyrus (SFG), from the same patients. The actual neuronal loss in PD is known to be greater in LSN compared to MSN, with the SFG relatively spared. Therefore significant probes (p≤0.01) with fold changes in LSN>MSN>SFG were defined as potentially “neuronal-loss-associated”. For example the tyrosine hydroxylase (TH) gene showed a fold change pattern of −14.5 (LSN), −4.9 (MSN) and no change (SFG). Any probes which were defined as “neuronal-loss-associated” and were not differentially expressed in residual laser-captured dopaminergic neurons (Cantuti-Castelvetri et al. study)[19] were removed from the subsequent analyses of all SN data sets.

**Online Database**

The differentially expressed gene list generated for each study by this re-analysis can be found with their respective p-values and fold changes can be found at http://ncsscr.griffith.edu.au/pdreview/2006/. A search can be performed individually using Entrez gene ID, gene symbol, or collectively by publicly available lists/pathways.

**Accession Numbers**

The National Center for Biotechnology Information Entrez Gene website (http://www.ncbi.nlm.nih.gov/sites/entrez?db = -gene) accession numbers (GeneIDs) for the genes named in the paper include: DDC(1644), DRD2(1813), SLC6A3(6531), TH(7054), IG2(3481).

**Supporting Information**

Table S1 Comparison of overlap in genes between PD-related transcriptomic studies. The enclosed table illustrates the increase in data convergence between PD-related transcriptomic studies following the implementation of our common analysis methodology. Found at: doi:10.1371/journal.pone.0004955.s001 (0.02 MB PDF)

Table S2 The differentially expressed probes generated by common analysis for each study. Ranked probe lists for each study generated by common analysis method with fold change and p-value. Found at: doi:10.1371/journal.pone.0004955.s002 (1.08 MB XLS)

Table S3 Over-represented pathway categories from IPA analysis. Over-represented pathway categories from IPA analysis of the differentially expressed probes of PD patients compared to controls: A- SN - Hauser study; B- SN - Zhang study; C- lateral SN - Moran study; D- medial SN - Moran study; E- SN - Lesnick study; F- Brodmann Area 9 - Zhang study; G- Putamen - Zhang study; H- Superior Frontal Gyrus - Moran study; I- Occipital Cortex - Vogt study; J- Putamen - Vogt study; K- Cerebellum - Vogt study; L- Whole Blood - Scherzer study; M- Laser Captured SN dopaminergic neurons - Castelvetri study. Found at: doi:10.1371/journal.pone.0004955.s003 (0.07 MB XLS)

Table S4 Selection of neuronal loss-associated genes. A work flow diagram and hypothetical examples illustrate the selection of neuronal loss-associated genes that were removed from the SN datasets prior to pathway analysis. Found at: doi:10.1371/journal.pone.0004955.s004 (0.12 MB PDF)

Table S3 ‘Neuronal-loss’ associated genes removed by correction paradigm. Identified probes with fold change’s for the three brain regions in the Moran study that followed the pattern, LSN>MSN>SFG. Found at: doi:10.1371/journal.pone.0004955.s005 (0.05 MB XLS)

Table S6 Identified probes with fold change’s for the three brain regions in the Moran study that followed the pattern, LSN>MSN>SFG. Over-represented pathway categories from
IPA analysis after neuronal correction of the differentially expressed probes from SN tissue: A- Hauser study; B- Zhang study; C- lateral SN; Moran study; D- medial SN Moran study; E- Lesnick study.

Found at: doi:10.1371/journal.pone.0004955.s006 (0.04 MB XLS)

Table S7  Fold changes in PD-related glial markers. The lack of differential expression of PD-related glial markers is illustrated in the enclosed table.

Found at: doi:10.1371/journal.pone.0004955.s007 (0.07 MB PDF)

References


Acknowledgments

The authors would like to acknowledge our PD colleagues who have kindly made their gene expression data available through public databases or forwarding it directly to us. We would like to thank Amanda Miotto and Othmar Korn for informatics support.

Author Contributions

Conceived and designed the experiments: GS NAM AC CGM. Performed the experiments: GS NAM. Analyzed the data: GS NAM AC MA CGM. Contributed reagents/materials/analysis tools: GS NAM AC MA PS AM CGM. Wrote the paper: GS NAM AC PS AM CGM.


