Citation for published version (APA):
Observations of extensive gene expression differences in the cerebellum and potential relevance to Alzheimer’s disease

Sally Chappell1, Tulsi Patel1, Tamar Guetta-Baranes1, Fei Sang2, Paul T. Francis3, Kevin Morgan1 and Keeley J. Brookes1*

Abstract

Objectives: In order to determine how gene expression is altered in disease it is of fundamental importance that the global distribution of gene expression levels across the disease-free brain are understood and how differences between tissue types might inform tissue choice for investigation of altered expression in disease state. The aim of this pilot project was to use RNA-sequencing to investigate gene expression differences between five general areas of post-mortem human brain (frontal, temporal, occipital, parietal and cerebellum), and in particular changes in gene expression in the cerebellum compared to cortex regions for genes relevant to Alzheimer’s disease, as the cerebellum is largely preserved from disease pathology and could be an area of interest for neuroprotective pathways.

Results: General gene expression profiles were found to be similar between cortical regions of the brain, however the cerebellum presented a distinct expression profile. Focused exploration of gene expression for genes associated with Alzheimer’s disease suggest that those involved in the immunity pathway show little expression in the brain. Furthermore some Alzheimer’s disease associated genes display significantly different expression in the cerebellum compared with other brain regions, which might indicate potential neuroprotective measures.

Keywords: RNA-sequence, Human brain, Cerebellum, Alzheimer’s disease

Introduction

RNA-sequencing is now at the forefront of genetic research into complex disease, for accurate quantitation of gene expression and has been found to provide more comprehensive view of the RNA landscape than microarrays [1]. Whereas most expression studies have focused on differential gene expression between control and disease tissue, few investigate differential expression observed between different tissues from the same individual. In order to determine how gene expression is altered in disease or pre-disease state it is of fundamental importance that the global distribution of gene expression levels is understood and how this might differ between tissue types to inform accurate tissue choice for investigation of altered expression in disease state. Online databases such as the Genotype-Tissue Expression (GTEx) project [2] are extremely helpful in determining this, however there is a need to gain insight for gene expression profiles in tissues from the same samples, as individual differences between samples (e.g. genetic background) might mask true positive findings and extenuate false positive finding.

The Brains for Dementia Research (BDR) cohort is an initiative to provide tissue for investigation into the aetiology of dementia. The cohort consists of well-defined clinically and neuropathological post-mortem brain tissue from healthy control and dementia samples, with the majority diagnosed as Alzheimer’s disease (AD). Recent RNA-sequencing studies comparing AD and control tissue compare either a single region, or a few specific regions or cell-types of the brain, producing variable results for the genes displaying significant difference in
expression between disease states [3–6]. This may be due to some genes showing expression differences between disease states in some regions but not others. Therefore in order to establish how reflective such studies are of each other, the gene expression profile across different brain regions must be conducted in corresponding tissue from the same subject. The discernment of which genes are differentially expressed between brain regions and not others could aid our understanding as to why certain regions of the brain are susceptible to various neuro-degenerative diseases and help inform which region should be studied in relation to certain genes or pathways.

**Main text**

**Methods**

Three neuro-pathologically confirmed healthy controls were selected from the BDR repository. With three biological replicates per brain region, there is 87% power to detect gene expression fold-changes > 2 [7]. Tissue samples from five regions of the brain (frontal, temporal, occipital, parietal and cerebellum) were provided for each of the three samples. Two of the samples were female, with one male sample, the average age of death was 71 years (±14 years), and average PMI = 50.3 h.

RNA was isolated from each of the regions using an in-house developed methodology preparing tissue with the Covaris cryoPrep system, to crush the tissue to increase the surface area and allowing efficient cell lysis in 1 ml TriZol following by RNA extraction with the RNAeasy Minikit from QIAGEN. A total of 2 μg of RNA was sent to the University of Nottingham DeepSeq Facility, for library preparation and sequencing. Total RNA samples were processed for ribosomal RNA depletion in order to enrich for non-coding and coding genes. Enriched RNA samples were then used to generate barcoded-sequencing libraries. Libraries were multiplexed onto 20 high output runs (2 × 75 bp) generating around 60 million reads per sample with sequencing performed using Illumina Next-seq500 platform.

The filtering pipeline was used to filter reads with low sequencing score as well as reads aligned to adaptor sequences. First, raw reads were trimmed against adaptors using ‘Sythe’ (https://github.com/vsbuffalo/scythe), then reads were quality trimmed using ‘Sickle’ (https://github.com/najoshi/sickle). Reads passing the filters were mapped onto the reference genome (hg19) in the context of known gene exon coordinates (Ensembl) using the ‘tophat’ mapping tool. (https://ccb.jhu.edu/software/tophat/index.shtml).

Read counts for each gene were calculated using ‘htseq-count’ (http://www-huber.embl.de/users/anders/HTSeq/doc/count.html). RPKM (reads per kilobase of transcript per million mapped reads) counts for all genes were calculated [8]. The RPKM is simply a normalized read count (stranded/sense reads) for a given gene. The read count of the exon-space of a gene is normalised against the total number of mapped reads against the total length of the gene’s exon-space. Genes with an average RPKM of < 1 were deemed non-discriminatory from background noise [9] and should be viewed with caution.

**Statistical analysis**

The programme ‘DESeq2’ [10] was used to detect differentially expressed genes between brain regions in a pair-wise fashion. This analysis uses the gene counts, and corrects for dispersion using Bayes theorem. The final values for differential expression are adjusted with a Benjamini-Hochberg (false discovery rate—FDR) with a P value < 0.05 deemed as significant.

**Results**

There was an average of 15,742 genes expressed across the 15 samples with an RPKM greater than 1, indicating gene expression above background noise [9]. Each region displayed a varying number of genes expressed, with the cerebellum expressing the greatest number of genes with an average of 16,576 across the three samples. The frontal, temporal, occipital and parietal cortex regions displayed expression of 15,450; 14,930; 15,995 and 15,757 genes, respectively. One-way ANOVA with post hoc Tukey revealed a significant difference between the number of genes expressed between the temporal and cerebellum regions (P = 0.023). No other comparisons were significant (P > 0.1). Across all regions the most highly expressed genes were SRP and RNU RNA genes, involved in translation and splicing mechanisms.

There was an average of 20,132 gene comparisons analysed per region (based on gene count), suggesting that around 1000 (5%) genes would be found to be differentially expressed by chance at the alpha significance threshold (P < 0.05 FDR corrected). In differential analyses between cortex regions far fewer genes were identified than this. Comparisons between the cerebellum and each cortex region suggested that 6–9 times more than the expected number of genes by chance were observed to be significantly differentially expressed (Table 1).

Gene expression in the cerebellum region was vastly different, with 11,770 unique genes differentially expressed compared to the other regions. Five thousand, three hundred and one genes (45%) were consistently differentially expressed between the cerebellum and the other cortex regions. The majority of these were concordant for direction of expression level change in the cerebellum compared to the other regions, with only 14 genes (0.3%) showing divergent expression direction changes between the cerebellum and each of the cortex regions.
Preliminary exploration of the data with Ingenuity’s Pathway Analysis (IPA) software (QIAGEN Bioinformatics) suggests that the common gene expression differences in the cerebellum indicate decreases in the development and quantity of neurons, and an increase in neuronal loss in the cerebellum. However it also suggests a decrease in long-term depression of the synapse and increases in long-term potentiation of cells, suggesting an increase in synaptic plasticity and therefore strength in the cerebellum.

Genes known to be involved in the familial early-onset form of AD, APP, PSEN1, and PSEN2, displayed varying levels of expression in the brain, with APP exhibiting high levels of expression, with PSEN1 and PSEN2 genes showing a lower level of expression (Fig. 1). Expression levels of APP and PSEN1 were observed to significantly lower in the cerebellum, whereas PSEN2 was higher.

The RNA-sequencing data for genes associated with the late-onset form of AD suggested considerable variation for the average level of gene expression (Fig. 1). Of particular note is that genes involved in immunity pathways were found to show very low levels of gene expression on average across all brain regions (RPKM ≤ 1), whilst those involved in cholesterol metabolism and endocytosis displayed moderate (RPKM > 15) to high gene (RPKM > 100) expression levels.

Nine of the late-onset form AD-associated genes displayed no significant changes in expression level between regions of the brain (APOE, SORL1, TREM2, ABCA7, ZCWPW1, HLA-DRB5, MS4A6A, CASS4, and TREML2). Three genes displayed a significantly higher in expression in the cerebellum compared to all other cortex regions (CELF1, CD2AP and EPHA1), whilst six genes displayed a significantly lower expression in the cerebellum (CLU, MEF2C, BIN1, FERMT2, SLC24A4 and INPP5D). Finally four genes displayed no significant change in expression between the regions of the brain except in one comparison with the cerebellum (Table 2).

The data generated here has a high level of concordance with data from GTEx version 6 (v6). The comparison of medium RPKM values between our data and that obtained for brain regions available in GTEx are highly correlated for frontal (Pearson r = 0.974 P < 0.001) and cerebellum regions (Pearson r = 0.966 P < 0.001). Data from GTEx also supports the data for direction change in expression levels between the frontal and cerebellum regions, with 17 genes showing distinguishable levels of RPKM change between the regions, 14 of them (82.4%) are concordant with the data generated here. Discordant genes were PTK2B, SORL1 and BIN1, showing the opposite direction of change in RPKM in the data provided here with that of GTEx v6.

### Discussion

The number of genes that exhibited expression differences between the different cortex regions of the brain were not above what was expected by chance, decreasing the credibility of those observations being true-positive findings. The key findings were the unique profile of the cerebellum gene expression, and the commonality of the genes differentially expressed between each of the four cortex comparisons. This echoes the findings by previous studies carried out using microarray technology [11–13]. These studies compared gene expression in the cerebellum to various other regions of the brain, and observed that while cortex regions had little variation of gene expression between them, when compared to the cerebellum over 1000 genes were found to be differentially expressed. The increased number of differentially expressed genes observed here, may reflect the greater accuracy and sensitivity of sequencing data over that of

<table>
<thead>
<tr>
<th></th>
<th>Frontal</th>
<th>Temporal</th>
<th>Occipital</th>
<th>Cerebellum</th>
<th>Parietal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal</td>
<td>2</td>
<td></td>
<td>385</td>
<td>9094</td>
<td>2</td>
</tr>
<tr>
<td>Occipital</td>
<td>274</td>
<td>227</td>
<td></td>
<td>6904</td>
<td>171</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>5658</td>
<td>5662</td>
<td>4437</td>
<td>8699</td>
<td></td>
</tr>
<tr>
<td>Parietal</td>
<td>13</td>
<td>1</td>
<td>77</td>
<td>5090</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Numbers of significantly differentially expressed genes identified by ‘DESeq2’ between brain regions (Benjamini-Hochberg adjusted significance level of < 0.05; top right corner), and numbers of significantly differentially expressed genes filtered by fold-change of greater than doubled or halved expression changes (bottom left corner)
microarray data. Tentatively, the most interesting observation was from the pathway analysis which suggests that the cerebellum has increased synaptic plasticity and strength, with the gene expression changes suggesting an increase in long-term potentiation of the cells, despite a decrease in the development and quantity of neurons. Synapse plasticity and therefore strength of cell–cell signalling is thought to underlie learning and memory [14].

A previous investigation of gene expression changes related to aging across different regions of the brain identified that whereas several regions of the brain displayed similar age-related gene expression changes to the frontal cortex, the cerebellum displays little correlation with these regions [15]. Their analysis suggests that the cerebellum shows fewer gene expression changes in relation to aging compared to other parts of the brain, which may account for the large number and commonality of gene differentially expressed between the cerebellum and cortex regions presented here. One of the suggestions by Fraser et al. [15] was that the cerebellum could be aging at a slower rate than other regions of the brain, and these differentially expressed genes might represent those associated with aging.

The cerebellum has long thought to be relatively preserved in AD [16, 17], and PET studies have utilized the cerebellum as a pseudo-control investigating neuro-inflammation due to the lack of difference shown for TSPO density between patients with AD and controls [18, 19]. This might possibly suggest that the cerebellum has some protective measures against the onset of aging and/or AD pathology. This however requires further exploration and could be the basis of gaining insight into the preservation of neurons in the brain and therefore therapeutic intervention.

This was further supported by the observation of genes that are purportedly associated with AD (familial and sporadic) via association studies display higher or lower expression levels in the cerebellum compared to cortex regions.
It was found that some genes known to be involved in familial and in the late-onset form of AD, were in fact expressed at low levels in the brain with some expressed at RPKM values below the cut-off that would indicate the levels could not be discriminated against background noise [9]. This was supported with expression data from GTEx v6. In particular, genes known to be associated with late-onset AD, involved in the immunity pathway, now a leading focus for therapeutic investigation for the disease, are expressed at very low levels in the brain. Data from GTEx suggests these genes involved in the immunity pathway are highly expressed in tissues such as whole blood, the lung and spleen.

Limitations
This pilot study lacks the power to discern real data on the variability between individuals. However it provides valuable data for gene expression across human brains regions for future reference. Ideally further RNA-sequencing of more specific brain regions on control and Alzheimer’s disease samples would add greater power to the study.

Abbreviations
BDR: Brains for Dementia Research; RPKM: reads per kilobase of transcript per million mapped reads; FDR: false rate discovery.

Authors’ contributions
SC—Conception of project, TP—Preparation of RNA, TGB—Laboratory work, FS—Bioinformatic analysis, PTF—Director of the BDR, acquisition of samples, choice of brain region and ethical approval, KM—Revision of manuscript, KJB—Laboratory work, analysis, construction of manuscript, project lead. All authors read and approved the final manuscript.

Author details
1 Human Genetics, School of Life Sciences, University of Nottingham, Nottingham, UK. 2 DeepSeq Facility, University of Nottingham, Nottingham, UK. 3 Wolfson Centre for Age Related Diseases, King’s College London, London, UK.

Acknowledgements
We would like to gratefully acknowledge all donors and their families for the tissue provided for this study. Human post-mortem tissue was obtained from the South West Dementia Brain Bank, London Neurodegenerative Diseases Brain Bank, Manchester Brain Bank, Newcastle Brain Tissue Resource and Oxford Brain Bank, members of the Brains for Dementia Research (BDR) Network. The BDR is jointly funded by Alzheimer’s Research UK and the Alzheimer’s Society in association with the Medical Research Council.

We also wish to acknowledge the neuropathologists at each centre and BDR Brain Bank-staff for the collection and classification of the samples. We thank the donor whose donation of brain tissue to the London Neurodegenerative Diseases Brain Bank allowed this work to take place. The Brain Bank is supported by the Medical Research Council and Brains for Dementia Research (jointly funded by the Alzheimer’s Society and Alzheimer’s Research UK).

The authors would like to thank the DeepSeq Facility, University of Nottingham for the sequencing and analysis of the RNA samples, in particular Sunir Malla for the library preparation and sequencing and Fei Sang for bioinformatics analysis.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
RNA sequencing data of the BDR samples will be publically available via Dementia Platform UK.

Consent to publish
Not applicable.

Table 2 Table of DESeq2 results for Alzheimer’s disease related genes, indicating fold-change of expression in index brain region in comparison to the second brain region, and the Benjamini-Hochberg FDR corrected significance observed for the change. Those changes deemed significant with P<0.05 are highlighted in red. Genes with no significant difference in expression level between any brain regions comparisons are shaded in grey.
Ethics approval and consent to participate
Brains for Dementia Research has ethics approval from London-City and East
NRES committee 08/H0704/128+5 and has deemed all approved requests for
tissue to have been approved by the committee.

Funding
This work was supported by an ARUK Network grant awarded to SC.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in pub-
lished maps and institutional affiliations.

Received: 23 July 2018   Accepted: 24 August 2018
Published online: 04 September 2018

References
1. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for tran-
2. Consortium, G. The Genotype-Tissue Expression (GTEx) project. Nat
reveal neurovascular defects, altered amyloid-beta homeostasis, and
4. Mastroeni D, et al. Laser-captured microglia in the Alzheimer’s and
Parkinson’s brain reveal unique regional expression profiles and suggest
2017;63:12–21.
5. Mills JD, et al. RNA-Seq analysis of the parietal cortex in Alzheimer’s
disease reveals alternatively spliced isoforms related to lipid metabolism.

of genes involved in immune response and mitochondrial processes in
10. Love MI, Huber W, Anders S. Moderated estimation of fold change and
11. Evans SJ, et al. DNA microarray analysis of functionally discrete human
12. Khaitovich P, et al. Regional patterns of gene expression in human and
14. Sweatt JD. Neural plasticity and behavior—sixty years of conceptual
Biol. 2005;3(9):e274.
16. Braak H, Braak E. Neuropathological stageing of Alzheimer-related
18. Kreisl WC, et al. In vivo radioligand binding to translocator protein cor-
19. Lyoo CH, et al. Cerebellum can serve as a pseudo-reference region in
Alzheimer disease to detect neuroinflammation measured with PET