Pathway Analysis of the Human Brain Transcriptome in Disease

Author:
Kavanagh, Tomas; Mills, James D; Kim, Woojin S; Halliday, Glenda; Janitz, Michael

Publication details:
Journal of Molecular Neuroscience
v. 51
Chapter No. 1
pp. 28-36
0895-8696 (ISSN)

Publication Date:
2013

Publisher DOI:
http://dx.doi.org/10.1007/s12031-012-9940-0

Pathway Analysis of the Human Brain Transcriptome in Disease.

Tomas Kavanagh[1], James D Mills[1], Woojin S Kim[2,3], Glenda M Halliday[2,3], Michael Janitz[1]*

[1]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW 2052, Australia

[2]Neuroscience Research Australia, Sydney, NSW 2031, Australia

[3]School of Medical Sciences, University of New South Wales, Sydney, NSW 2052, Australia

[*]**Corresponding Author***:* School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW 2052, Australia, Tel.: +612 938 58608, Fax: +612 938 51483, Email: m.janitz@unsw.edu.au

Abstract

Pathway analysis is a powerful method for discerning differentially regulated genes and elucidating their biological importance. It allows for the identification of perturbed or aberrantly expressed genes within a biological context from extensive data sets and offers a simplistic approach for interrogating such datasets. With the growing use of microarrays and RNA-Seq data for genome wide studies is growing at an alarming rate and the use of deep sequencing is revealing elements of the genome previously uncharacterised. Through the employment of pathway analysis, mechanisms in complex diseases may be explored, and novel causatives found primarily through differentially regulated genes. Further, with the implementation of next generation sequencing (NGS) a deeper resolution may be attained, particularly in identification of isoform diversity and SNP's. Here we look at a broad overview of pathway analysis in the human brain transcriptome and its relevance in teasing out underlying causes of complex diseases. We will outline processes in data gathering and analysis of particular diseases in which these approaches have been successful.

Keywords: transcriptome; pathway analysis; human brain; brain disorders; gene expression; RNA-Seq

Introduction

In the complex transcriptional environment of the brain, identification of differentially regulated genes is a significant challenge. This is especially true in cases of neurodegenerative diseases in which ageing itself is causative of many of the changes seen (Lu et al. 2004). To complicate the identification of causative genes in neurodegenerative diseases further; the brain has an enormous range of transcript isoforms, and many differentially expressed isoforms have been linked to neurodegenerative diseases (Courtney et al. 2010; Faustino and Cooper 2003; Mills and Janitz 2012; Tollervey et al. 2011; Twine et al. 2011). These isoforms may be differentially expressed across tissues or in disease, or alternative splicing and erroneous splicing may have an impact on disease (Mills and Janitz 2012; Tollervey et al. 2011; Twine et al. 2011). It is imperative that a deep understanding of the brain transcriptome be developed so that diseases of the brain might be better treated.

In identifying key differences in complex disease states systems biology looks to identify differentially regulated genes. Whilst there are many potential methods to study the transcriptome, the use of microarrays is likely the most prevalent. However, microarray analysis is not without faults and the recent improvement in NGS and development of RNA-Seq - which has greater reproducibility, resolution and cost effectiveness when analysing an entire transcriptome – looks set to overtake microarrays (Shendure 2008).

As the wealth of data generated by the aforementioned techniques grows, the interrogation of the datasets for biologically relevant information becomes more complex. Traditionally these approaches provide a list of genes identified as differentially expressed. Pathway Analysis has gained favour for its ability to reduce these lists to a series of ‘pathways’ in biological systems. These pathways are formed of gene and protein products within a defined context of biology, e.g. sphingolipid metabolism – Human. Examples of

these pathways include those formed from Gene Ontology annotated genes drawn from public databases (Emmert-Streib and Glazko 2011; Khatri et al. 2012). The analysis may be implemented in any number of programs or databases and include online resources such as The **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (DAVID) (Huang et al. 2009) or MetaCore (www.genego.com).

Use of DAVID requires little technical expertise yet can determine the presence of affected pathways and allow for simplified targeted identification and analysis of aberrant genes. To complete a DAVID analysis it is necessary to have a sufficiently extensive gene list ranging from hundreds to thousands of genes and each gene needs to have been selected by statistical analysis (e.g. *t-test*) and show significant fold change in expression (Huang et al. 2009). DAVID's output of affected gene pathways may then be used in an exploratory fashion to find differentially expressed genes of biological relevance to the state being questioned.

The use of pathway analysis has seen extensive use throughout microarray and RNA-Seq analysis. These experiments have ranged from aiming to generate a generalised understanding of gene expression changes or differences during development or ageing and between tissue types (Chen et al. 2011; Lerch et al. 2012) or in disease (Sutherland et al. 2011). Such studies have encompassed Alzheimer's disease (AD), schizophrenia (SCZ), autism spectrum disease (ASD), which will be discussed here, amongst others.

Pathway analysis offers a powerful analytical tool for studying the transcriptome within the human brain. It has the ability to readily identify differentially expressed genes, and also to isolate them into biological groupings based on functional relation, structural relation or position in a cell amongst other grouping. In doing so, pathway analysis reduces the complexity of such tasks involving large gene lists of differentially regulated genes and

hance has strong applications in studies of the transcriptome using either RNA-Seq or microarray for data collection.

Data Collection

While pathway analysis is applicable to many data sets, including metabolomics, proteomics and gene expression (Emmert-Streib and Glazko 2011) its use in interpreting transcriptome data generated by microarray or RNA-Seq is of particular interest here.

Microarray

Microarrays have become a popular method for elucidating gene expression patterns in disease states. Microarrays use islands of oligo-primer probes fixed to a solid surface such as glass to hybridise to cDNA samples generated from RNA (Brown and Botstein 1999). By setting up these probes in an array, relative expression of the library of genes being screened can be visualised when labelled sample is washed across the array for hybridisation. Appropriate use and placement of probes within genes increases the specificity of this process and can provide sequence information about the expression of transcripts within the sample being analysed. Whilst microarrays improve with increasing knowledge of genomes they still rely on a reference genomic sequence, and as such are incapable of de novo sequence identification (Kapur et al. 2008; Shendure 2008).

Standard cDNA arrays allow for the analysis of the absolute, often for the most common isoforms, expression differential to be measured. Microarrays produce extensive gene lists showing differential regulation of those genes between the states being investigated. However, the basis of hybridisation of probe to cDNA strand is by no means flawless. As a result microarray data needs to be corrected for hybridisation background. For specialised study of isoforms in transcriptomics, specialised microarrays, such as exon and tiling arrays,

have been developed. Exon arrays are rich in probes specific for exons of transcripts based on libraries of annotated, partially supported or estimated exons. Each probe set will target one exon and show expression levels of each exon type providing an accurate representation of gene expression and suggesting the likely composition of transcripts (Kapur et al. 2008).

Many computing technologies have been developed for analysis of microarray data. These include GeneBASE a tool for estimating expression levels of genes [*www.stanford.edu/group/wonglab/GeneBASE/*] and microarray analysis of differential splicing (MADS), an improved program for transcript isoform construction from exon tiling arrays (Xing et al. 2008). These, and other programs, allow for more accurate processing of microarray data by removing hybridisation background, more accurately quantifying gene expression and assembling RNA isoforms. This in turn allows for deeper analysis of the transcriptome via microarray.

Despite the usefulness of microarrays in identifying deregulated genes and even transcripts, significant flaws in microarray technology are becoming apparent. Non-specific signals and cross-hybridisation are significant and can cause loss of reproducibility among microarray experiments (Homer et al. 2008). The reliance on hybridisation, analogue signatures and dependence on availability of the reference genome sequence make it therefore difficult to detect rare, low abundance or structurally similar isoforms (Chen et al. 1997; Homer et al. 2008). This makes high resolution studies of the transcriptome problematic.

RNA-Seq

RNA-Seq is a recent technology that utilises NGS platforms to sequence short cDNA fragments of around 100bp (Costa et al. 2010; Huang et al. 2011; Wang et al. 2009) and thus gain a snapshot of the human transcriptome at a particular point in time. This provides a view

of not only what genes are being expressed in a particular tissue and to what degree, but will also reveal SNPs genotype within transcribed parts of the genome. Moreover alternatively spliced isoform type, abundance, diversity and regulation may be determined (Fang and Cui 2011; Lerch et al. 2012; Trapnell et al. 2009). It is also possible to detect RNA editing and coding/non-coding RNA's and potential sites of action for non-coding RNA, revealing the complex nature of transcriptome dynamics (Peng et al. 2012). RNA-Seq overcomes the shortcomings of microarray, that, for many organisms, rely on reference genomes that are incomplete for the entire transcriptome and thus fail to identify any new or unknown transcript isoforms (Ameur et al. 2011; Martin and Wang 2011; Wang et al. 2009).

RNA-Seq offers base pair resolution and strong coverage of the transcriptome. This allows for the determination of transcript structure even as the result of alternative splicing (AS) or miss-splicing. The ability to determine expression levels of individual exons is, perhaps, RNA-Seqs most useful feature. Using this technology it has been estimated that approximately 95% of multi-exon genes undergo AS to give potentially, a hundred thousand splicing events occurring in major human tissues (Pan et al. 2008). However, the data quantity produced by RNA-Seq is overwhelming since the transcriptomic analysis of one human individual can result in ~70 Gbp of sequence reads (Peng et al. 2012).

In designing RNA-Seq experiments, issues such as read coverage or sample size are key considerations. Most of the transcripts are expressed at relatively low levels and detection of their expression will require generation of at last 40 million reads per sample (Toung et al. 2011). Further it is essential to include sufficient number of biological replicates and of technical replicates to ensure reproducibility and facilitate quantification of individual genes of interest using qPCR (Fang and Cui 2011). For example current version of Cufflinks, a transcript assembler, requires sequence input from at least three biological replicates to

provide sufficient statistical power for calculation of variability in gene expression between samples.

With the growing popularity of RNA-Seq in genome wide association studies, many new computing and analysis methods have been developed to deal with the influx of data. For sequence alignment one such program includes Tophat which when extended to Cufflinks is not only useful in aligning sequences but can reconstruct isoform structure and align these to the genome (Trapnell et al. 2012). This allows for deeper analysis of differentially expressed genes, with not only the absolute expression of a gene seen, but also the expression of its isoforms. This enhanced resolution increases RNA-Seq's potential to elucidate mechanisms of poorly understood complex disease processes.

Pathway analysis

When using pathway analysis to explore RNA-Seq or microarray data for differentially expressed genes, there is a number of options available.

Over-representation analysis was the first type of approach developed to take advantage of microarray and Gene Ontology. This method selects genes from the total list of genes generated by RNA-Seq or microarray, often by fold change cut off (commonly two fold change) with a p-value of 0.05 (Emmert-Streib and Glazko 2011; Khatri et al. 2012). These genes are used as input in analysis packages such as DAVID or MetaCore which aligns these genes to database knowledge (Huang et al. 2009). Pathways will be ordered according to overrepresentation, i.e. the number of genes differentially expressed from the number of genes defined by the pathway. Each pathway may then be explored for differentially expressed genes within biological context; this represents a knowledge based identification system of differentially expressed gene candidates for further study.

Flaws in the overrepresentation method include a loss of data as selection criteria, such as fold-change and p-value cut-off for reducing gene lists generated by RNA-Seq or microarray, are often arbitrary. They may ignore genes that fall just shy of these criteria or are not sufficiently affected by the state under investigation, such as disease, when small disturbances may lead to crucial loss of pathway homeostasis further downstream (Subramanian et al. 2005).

To compensate for the weaknesses of over-representation analysis two other major methods were developed. Functional class scoring uses a multitude of statistics applied to gene level changes recorded in experiments. These results are carried through to the pathway level to aid in assessing the significance of an affected pathway (Ackermann and Strimmer 2009). The disadvantages of this method lie in its treatment of pathways, as in over-representation analysis, as individual pathways when many have overlapping genes and functions. Further use of ranking at pathway level can lead to genes with a high degree of differential expression which may be considered equal to those with a lower degree of differential expression (Pavlidis et al. 2004).

The final method, pathway topology analysis, utilises further information from specialised databases (e.g. KEGG) such as metabolite production or protein sub cellular location to assess how much one pathway may be affected by differentially expressed genes (Kanehisa and Goto 2000). However, pathway topology changes from cell to cell, and as such a poor or no characterisation and annotation for some cell lines is available which presents a major drawback in this method (Green and Karp 2006; Bauer-Mehren et al. 2009). Furthermore the method does not fully compensate for the interdependencies between similar pathways.

Whilst a myriad of software packages exist for performing pathway analysis on large gene sets - including DAVID, MetaCore, GenGen, EASE and many others – MetaCore, a professional software package and DAVID, an open source package, are the main focus here. Whilst MetaCore and DAVID both perform overrepresentation analyses they utilise different databases. Where MetaCore relies on its own manually curated database, DAVID utilises the BioCarta and KEGG databases amongst many others (Huang et al. 2009, ***david****.abcc.ncifcrf.gov/*). This gives DAVID a unique flexibility, allowing the user to view gene lists in varying biological perspectives. However, MetaCore offers many integrate functions including algorithms to enable Pathway Topology analysis, pathway construction with integrated information on protein-protein interactions, mass spectrometry data, nuclear magnetic resonance data and metabolomics databases and a strong drug design and analysis amongst many other functions (www.genego.com).

Both programmes have simple setup and analyse wizards to provide guided usage and provide access to many reference databases such as Ensemble and UCSC Genome browser. MetaCore features nicer design of pathways and integrates experimental values. This combined with MetaCores simple storage and activation of multiple gene lists allows the user to directly compare experiments with visual plotting of expression differences. DAVID features a gene ID converter to ensure the majority of genes from a list are incorporated into the analysis but can only accept 3000 genes in a single job (Huang et al. 2009, ***david****.abcc.ncifcrf.gov/*). Each Pathway Analysis programme offers its own unique features and failings. As such due consideration should be given to each before a decision is made as to which programme to favour, with researchers often choosing multiple programmes to increase reproducibility and robustness of their results.

Choosing the appropriate method of analysis is key for getting the best data out of gene lists (Breitling et al. 2004). This however may be affected by data collection, and

database choices for analysis depending on previous knowledge of pathways in a cell line or phenotype of interest (Green and Karp 2006). This is also dependent on the sort of control implemented, whether this is a self-contained, comparison of differential expression between two phenotypes, or a competitive, differential expression of genes compared to a control set in the same list.

Application of pathway analysis in disease studies

Diseases of the brain are often complex, with gradual progressive changes. Many neurodegenerative diseases such as AD have greater risk with age, and once the process is initiated it advances with age. It is vital that an understanding of the key mechanisms of complex diseases in the brain be understood, particularly at early stages so that preventative measures might be taken to halt disease progression.

Alzheimer's Disease (AD)

AD is the most prevalent form of dementia (~60% of dementia) and is set to grow rapidly in the western world as population's age (Lobo et al. 2000; Fratiglioni et al. 1999). AD is a complex disease and whilst it has been associated with a build-up of amyloid beta (Aβ) and Tau proteins the parthenogenesis of the disease is still largely unknown (Bertram et al. 2010). The use of transcriptomics and pathway analysis of gene expression within human brain tissue offers a powerful method of elucidating aberrant expression of genes that may be causing or compounding the disease.

~~ATP-binding cassette transporter A7 (ABCA7) has also been recently identified as a strong candidate gene for late-onset AD (Hollingworth et al. 2011). ABCA7 is a member of the ABCA subfamily which is characterized by the ability to transport lipids across membranes (Kim et al. 2008). ABCA7 is expressed in the brain and has been shown to~~

~~suppress the production of the neurotoxic amyloid-beta peptide, a key pathological process in AD (Chan et al. 2008). Two isoforms of ABCA7, arising from alternative splicing, have been identified in the human brain (Ikeda et al. 2003). Type I ABCA7 splice variant has been detected on the cell surface and in intracellular compartments, whereas Type II appears to remain in the endoplasmic reticulum (Ikeda et al. 2003). Different variants of this protein indicate possibly different biological functions. Apart from the lipid transport function, two groups have identified a role for ABCA7 in phagocytosis (Iwamoto et al. 2006; Jehle et al. 2006). They have demonstrated that macrophage ABCA7 plays a key role in phagocytosis of apoptotic debris. Since ABCA7 is strongly expressed in human microglia (Kim et al. 2006) and the fact that microglia actively phagocytose apoptotic debris in the CNS (Stolzing and Grune 2004), it is plausible that microglial ABCA7 could be involved in phagocytosis of apoptotic debris arising from neurodegenerative disease.~~

Increasingly it is being suggested that microRNA's (miRNA's) and transcriptional regulation play a key role in the development of spontaneous AD. As these miRNA's are theorised to control expression of genes and direct splicing and regulation of mRNA transcripts they play a key role in the development of neurodegenerative diseases (Satoh 2012). It has been seen in many studies that miRNAs is increasingly miss regulated. This includes the miss regulation of miR-146a which is controlled by NF-κB and is associated with sustained inflammation seen in AD brains (Li et al. 2011). Further groups have shown miRNA's that target BACE1-regulated formation of Aβ are significantly reduced in AD frontal lobes (Kellett and Hooper 2009).

Studies into alternative splicing of genes previously correlated to AD pathology have shown differential expression of transcript isoforms, a potential mediator of disease. RNA-Seq data of AD brains analysed in DAVID shows differential expression of apolipoprotein E

(*APOE*) isoforms, with *APOE-001* and *APOE-002* showing decreased expression in AD brains compared to normal brains, and *APOE-005* showing significant up-regulation in AD brains compared to normal brains (Twine et al. 2011). Many other cases of aberrant splicing have been linked to AD pathology including cryptic and deletion splicing of presenilin 1 (*PSEN1*) leading to loss or cryptic inclusion of exon 4. Altered splicing has been suggested to affect amyloid precursor protein (*APP*) distribution in the AD brain and mutations in other proteins such as progranulin (*GRN*) (an anti-inflammatory protein), prevent appropriate splicing and may lead to a loss of neuro-protection (Golde et al. 1990; Cruts et al. 2006).

Meta-analysis of late onset AD cases has shown a number of variants in the bridging integrator 1 (*BIN1*), complement component (3b/4b) receptor 1 (*CR1*) and clusterin (*CLU*) genes to be involved in AD pathogenesis (Lambert et al. 2009; Harold et al. 2009). However the *CLU* gene has not been replicated in some studies utilising GenGen: Genetic Genomics analysis of complex Data in meta-analysis of Alzheimer's expression data (Hu et al. 2011). The same authors also identified phosphatidylinositol binding clathrin assembly protein (*PICALM*) variants, yet the presence of *PICALM* was seen as insignificant when corrected for *APOE* variant presence, which is recognised as the most significant genetic marker for AD. *PICALM* is involved in clathrin-mediated endocytosis of *APP* fragments before cleavage by γ-secretase to yield Aβ, and as such it has been suggested that dysfunctional *PICALM* may hinder Aβ clearance (Bertram et al. 2010). The discovery of the *CR1* locus in this study shows a relationship to inflammation that the authors suggest compounds environmental factors such as diabetes and high blood pressure that appear to affect the development of dementia.

ATP-binding cassette transporter A7 (ABCA7) has also been recently identified as a strong candidate gene for late-onset AD (Hollingworth et al. 2011). ABCA7 is a member of the ABCA subfamily which is characterized by the ability to transport lipids across

membranes (Kim et al. 2008). ABCA7 is expressed in the brain and has been shown to suppress the production of the neurotoxic amyloid-beta peptide, a key pathological process in AD (Chan et al. 2008). Two isoforms of ABCA7, arising from alternative splicing, have been identified in the human brain (Ikeda et al. 2003). Type I ABCA7 splice variant has been detected on the cell surface and in intracellular compartments, whereas Type II appears to remain in the endoplasmic reticulum (Ikeda et al. 2003). Different variants of this protein indicate possibly different biological functions. Apart from the lipid transport function, two groups have identified a role for ABCA7 in phagocytosis (Iwamoto et al. 2006; Jehle et al. 2006). They have demonstrated that macrophage ABCA7 plays a key role in phagocytosis of apoptotic debris. Since ABCA7 is strongly expressed in human microglia (Kim et al. 2006) and the fact that microglia actively phagocytose apoptotic debris in the CNS (Stolzing and Grune 2004), it is plausible that microglial ABCA7 could be involved in phagocytosis of apoptotic debris arising from neurodegenerative disease.

Correlation of microarray data using Expression Analysis Systematic Explorer (EASE) has shown that incipient AD brains have a curiously high tumour suppressor transcription factor expression ratio (Blalock et al. 2004). These tumour suppressors, such as those from the *TGF-β* family and cyclin-dependent kinase 2 associated protein 1 (*CDK2AP1*), were shown to be involved in apoptosis and cell cycle amongst other functions, whilst down-regulated tumour suppressor genes were those related to growth and proliferation, such as v-myc myelocytomatosis viral oncogene homolog (*MYC*) and transcription factor Dp-1 (*DP1*). Whilst many of these tumour suppressors are involved in apoptotic pathways many have alternative roles, such as neuron senescence and extension, or re-entry into cell cycle (Blalock et al. 2004). As such differential expression of tumour suppressor genes may have complex effects on cells within the AD brain, including cell cycle re-entry, increased

myelination, cessation of protein biosynthesis and not just apoptosis. This varies with cell type and position within the AD brain.

Analysis performed within DAVID on synaptically localised mRNA showed an increase in neuroplasticity genes such as glutamate receptor, ionotropic, AMPA 2 (*GR1A2*) and solute carrier family 1, member 2 (*SLC1A2*). This suggests activation of inhibitory pathways as these proteins would hold synapses in a low conductance state. Further an increase in 3'UTR control sequences that regulate translation in the synapse were observed in many of these up-regulated neuro-plasticity genes (Williams et al. 2009).

A study on micro-vessels in the AD brain showed more than 2000 genes aredifferentially expressed compared to normal brains with a large number of these genes mapping to inflammatory and immune responses, signal transduction and neural development in the GeneSpringGX programme (Wang et al. 2012). Whilst these processes may not play pivotal roles in AD progression they may compound the effects of the disease. In particular over expression of cytokines and their receptors such as interleukin 1 receptor, type II (*IL1R2*) may significantly affect cell activity and be related to the decrease in gene expression of signal transduction related genes.

A major problem with studying the brain is the significant difference in architecture and functionality across regions. As such studies need to decide whether to take whole brain homogenates, single cell samples or specifically grey or white matter samples. Analysis of dissected grey and white matter samples through DAVID allowed for the identification of AD processes such as down-regulation stabilisation of ryanodine-sensitive Ca^{2+} that may have been masked in previous studies by white matter expression profiles (Blalock et al. 2011).

In order to effectively treat AD early disease detection and neuroprotective treatments are essential. To accomplish this, diagnostic indicators need to be differentiated from processes that occur with aging as many of these processes overlap (Blalock et al. 2004). As such it is useful to have animal and single cell models to investigate incipient AD. Mice (5XFAD) have been used to show that early stage AD shows particular patterns of gene expression. This included up-regulation of solute carrier family 18 member 3 (*SLC18A3*) and cholinergic receptor, muscarine 1(*CHRM1*) and the down-regulation of genes such as caveolin 3 (*CAV3*) and purkinje cell protein 2 (*PCP2*) as revealed in combinatorial use of EASE, Ingenuity Pathway Analysis (IPA) and GeneSpringGX packages (Kim et al. 2012). Such studies may aid in finding markers for incipient AD that will allow early detection so neuro-protective treatments might be implemented.

Pathway analysis of AD brain tissue has indicated thousands of differentially expressed genes to be involved in the disease progression. Whilst the number of genes is daunting these studies have significantly improved our knowledge of AD pathology and increase the potential for treatments to be developed. Whilst the complex nature of the disease may prevent implementation of a single therapeutic drug, a system of treatments could be developed with sufficient knowledge of pathway disruption in AD.

Autism Spectrum Disease (ASD)

ASD is a heterogeneous collection of clinical conditions that are highly heritable with recurrence in siblings tens of times higher than the general population (Ziats and Rennert 2011; Voineagu et al. 2011). Attempts to unravel a genetic cause for the disease have largely been unsuccessful. As the disease is highly complex, representing a broad clinical presentation that is likely affected by a series of genetic, epigenetic and environmental factors.

However, recent implementation of RNA-Seq and pathway analysis has begun to elucidate key changes in the autistic brain previously unknown.

A study on induced pluripotent stem cells (iPSCs) derived from a number of patients and control groups, showed that a series of genes were dysregulated during neuronal differentiation. Further it was seen that pseudogenes, lincRNAs and other regulatory RNAs were dysregulated during neurogenesis in ASD. These changes, revealed by DAVID, were seen to affect regulation of some gene candidates for ASD and included chromatin remodelers, cell adhesion molecules and regulatory RNA species for HOX boxes and other developmental gene sets (Lin et al. 2011). Changes in these genes likely contribute to the lack of definition observed between compartments in ASD brains.

It has been observed that transcriptome of the ASD brain lacks definition when compared to the normal brain. This has been investigated in the temporal and frontal lobes of ASD and normal patients, it was shown that the frontal and temporal lobes of ASD brains lack the ultrastructural organisation observed in normal brains (Voineagu et al. 2011). Ataxin 2-bonding protein 1 (*A2BP1*), a splicing factor specific to neural and muscle cells has been highlighted as differentiating ASD brains. Through the use of RNA-Seq and analysis in DAVID it has been shown that down regulation of the *A2BP1* leads to an increase in differential splicing and miss splicing of several genes in ASD brains when compared to control brains. Such targets included neuronal cell adhesion molecule (*NRCAM*) and glutamate receptor, ionotropicm N-methyl D-aspartate 1 (*GRIN1*), involved in synapse formation, amongst other genes (Voineagu et al. 2011).

Pathway analysis of autism affected brains has revealed that ASD perturbations may affect central cytokine signalling pathways. This is seen in network hubs, from IPAs Pathway tool, highly enriched for *Tnf, NF-κB, TgfB1, Myc, Jnk* and *Mapk* (Ziats and Rennert 2011).

The over-representation of many of these hubs will have significant downstream effects in many pathways and, the effects of these need to be explored in the disease context. The patterning of these network hubs and cell-type specific expression led the authors to suggest that glial cells as well as neurons are also affected in ASD brains.

Knowledge of the genetic background of ASD was previously limited, but with implementation of high-throughput systems biology approaches our understanding of the disease may rapidly progress.

Schizophrenia (SCZ)

The molecular aberrations leading to SCZ and bi-polar disorder (BP) are largely unknown though recent attempts at pathway analysis of data generated by high throughput techniques such as proteomics, microarray and RNA-Seq attempt to address this. Whilst SCZ and BP are clinically distinct disorders they share a number of genetic risk factors (Lin et al. 2011). A wide range of differentially expressed genes have been observed in common in these two diseases through statistical analysis and use of EASE, DAVID, GenMAPP and MAPPFinder programmes. This includes genes such as neurexin 3 (*NRXN3*), glutamate decarboxylase 1 (*GAD1*), claudin 11 (*CLDN11*) amongst many others (Konradi et al. 2004; Hashimoto et al. 2008; Dracheva et al. 2006; Chu et al. 2009). These genes are primarily related to synapse formation and myelin synthesis which play key roles in brain development and diseases of the brain.

SCZ data-mining in the Stanley Medical Research Institutes (SMRI) database and qPCR has identified an increase in apoptotic signals, particularly Tumour Necrosis Factor Superfamily Member 6 (FAS) receptor and Tumour Necrosis Factor [ligand] Superfamily member 13 (*TNFSF13*) (Catts and Weickert 2012). However it is noted that SCZ brains often have increased neural density with decreased size, yet a loss of dendritic field is observed.

This suggests that the role of these apoptotic factors in SCZ is not leading to cell death, but likely affect dendrite regression amongst other subtle effects on cellular components leading to interneuronal signalling deficits (Catts and Weickert 2012). In contrast, increases in TNFSF13 and FAS receptor RNAs were not observed in BP patients.

As expression profiles are explored between these two diseases a number of discriminating diagnostic markers may be identified, allowing for rapid determination of disease type without the need of psychoanalysis.

Concluding Remarks

Pathway analysis of high-throughput techniques, such as microarray and RNA-Seq, offers the chance to unravel molecular causes of complex diseases as discussed above. With increasing knowledge of complex diseases the probability of early detection, prevention or treatment development for complex diseases, such as AD, increases. Whilst many challenges still oppose progress in the field, such as data handling and storage, analysis of complex and lengthy gene lists and integration of this knowledge into complex diseases, the improvement and application of these technologies will significantly increase our understanding of complex polygenic diseases and provide novel targets for further analysis.

Acknowledgements

The work was supported by the National Health and Medical Research Council of Australia (1022325 to WSK, 630434 to GMH).

References

Ackermann M, Strimmer K (2009). A general modular framework for gene set enrichment analysis. BMC Bioinformatics. 10**,** 47.

Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L, Feuk L (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat Struct Mol Biol. 18**,** 1435-1440.

Bauer-Mehren A, Furlong LI, Sanz F (2009). Pathway databases and tools for their exploitation: benefits, current limitations and challenges. Mol Syst Biol. 5**,** 290.

Bertram L, Lill CM, Tanzi RE (2010). The genetics of Alzheimer disease: back to the future. Neuron. 68**,** 270-281.

Blalock EM, Buechel HM, Popovic J, Geddes JW, Landfield PW (2011). Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. J Chem Neuroanat. 42**,** 118-126.

Blalock EM, Geddes JW, Chen KC, Porter NM, Markesbery WR, Landfield PW (2004). Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. Proc Natl Acad Sci U S A. 101**,** 2173-2178.

Breitling R, Amtmann A, Herzyk P (2004). Iterative Group Analysis (iGA): a simple tool to enhance sensitivity and facilitate interpretation of microarray experiments. BMC Bioinformatics. 5**,** 34.

Brown PO, Botstein D (1999). Exploring the new world of the genome with DNA microarrays. Nat Genet. 21**,** 33-37.

Catts VS, Weickert CS (2012). Gene expression analysis implicates a death receptor pathway in schizophrenia pathology. PLoS One. 7**,** e35511.

Chan SL, Kim WS, Kwok JB, Hill AF, Cappai R, Rye KA, Garner B (2008). ATP-binding cassette transporter A7 regulates processing of amyloid precursor protein in vitro. J Neurochem. 106, 793-804.

Chen G, Yin K, Shi L, Fang Y, Qi Y, Li P, Luo J, He B, Liu M, Shi T (2011). Comparative analysis of human protein-coding and noncoding RNAs between brain and 10 mixed cell lines by RNA-Seq. PLoS One. 6**,** e28318.

Chen Y, Dougherty ER, Bittner ML (1997). Ratio-based decisions and the quantitative analysis of cDNA microarray images. J Biomed Opt. 2**,** 364-374.

Chu TT, Liu Y, Kemether E (2009). Thalamic transcriptome screening in three psychiatric states. J Hum Genet. 54**,** 665-675.

Costa V, Angelini C, De Feis I, Ciccodicola A (2010). Uncovering the complexity of transcriptomes with RNA-Seq. J Biomed Biotechnol. 2010**,** 853916.

Courtney E, Kornfeld S, Janitz K, Janitz M (2010). Transcriptome profiling in neurodegenerative disease. J Neurosci Methods. 193**,** 189-202.

Cruts M, Gijselinck I, van der Zee J, Engelborghs S, Wils H, Pirici D, Rademakers R, Vandenberghe R, Dermaut B, Martin JJ, van Duijn C, Peeters K, Sciot R, Santens P, De Pooter T, Mattheijssens M, Van den Broeck M, Cuijt I, Vennekens K, De Deyn PP, Kumar-Singh S, Van Broeckhoven C (2006). Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. Nature. 442**,** 920-924.

Dracheva S, Davis KL, Chin B, Woo DA, Schmeidler J, Haroutunian V (2006). Myelin-associated mRNA and protein expression deficits in the anterior cingulate cortex and hippocampus in elderly schizophrenia patients. Neurobiol Dis. 21**,** 531-540.

Emmert-Streib F, Glazko GV (2011). Pathway analysis of expression data: deciphering functional building blocks of complex diseases. PLoS Comput Biol. 7**,** e1002053.

Fang Z, Cui X (2011). Design and validation issues in RNA-Seq experiments. Brief Bioinform. 12**,** 280-287.

Faustino NA, Cooper TA (2003). Pre-mRNA splicing and human disease. Genes Dev. 17**,** 419-437.

Fratiglioni L, De Ronchi D, Aguero-Torres H (1999). Worldwide prevalence and incidence of dementia. Drugs Aging. 15**,** 365-375.

Golde TE, Estus S, Usiak M, Younkin LH, Younkin SG (1990). Expression of beta amyloid protein precursor mRNAs: recognition of a novel alternatively spliced form and quantitation in Alzheimer's disease using PCR. Neuron. 4**,** 253-267.

Green ML, Karp PD (2006). The outcomes of pathway database computations depend on pathway ontology. Nucleic Acids Res. 34**,** 3687-3697.

Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Williams A, Jones N, Thomas C, Stretton A, Morgan AR, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Morgan K, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Holmes C, Mann D, Smith AD, Love S, Kehoe PG, Hardy J, Mead S, Fox N, Rossor M, Collinge J, Maier W, Jessen F, Schurmann B, van den Bussche H, Heuser I, Kornhuber J, Wiltfang J, Dichgans M, Frolich L, Hampel H, Hull M, Rujescu D, Goate AM, Kauwe JS, Cruchaga C, Nowotny P, Morris JC, Mayo K, Sleegers K, Bettens K, Engelborghs S, De Deyn PP, Van Broeckhoven C, Livingston G, Bass NJ, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Al-Chalabi A, Shaw CE, Tsolaki M, Singleton AB, Guerreiro R, Muhleisen TW, Nothen MM, Moebus S, Jockel KH, Klopp N, Wichmann HE, Carrasquillo MM, Pankratz VS, Younkin SG, Holmans PA, O'Donovan M, Owen MJ, Williams J (2009). Genome-wide association study

identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet. 41**,** 1088-1093.

Hashimoto T, Arion D, Unger T, Maldonado-Aviles JG, Morris HM, Volk DW, Mirnics K, Lewis DA (2008). Alterations in GABA-related transcriptome in the dorsolateral prefrontal cortex of subjects with schizophrenia. Mol Psychiatry. 13**,** 147-161.

Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, Abraham R, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Jones N, Stretton A, Thomas C, Richards A, Ivanov D, Widdowson C, Chapman J, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Holmes C, Mann D, Smith AD, Beaumont H, Warden D, Wilcock G, Love S, Kehoe PG, Hooper NM, Vardy ER, Hardy J, Mead S, Fox NC, Rossor M, Collinge J, Maier W, Jessen F, Rüther E, Schürmann B, Heun R, Kölsch H, van den Bussche H, Heuser I, Kornhuber J, Wiltfang J, Dichgans M, Frölich L, Hampel H, Gallacher J, Hüll M, Rujescu D, Giegling I, Goate AM, Kauwe JS, Cruchaga C, Nowotny P, Morris JC, Mayo K, Sleegers K, Bettens K, Engelborghs S, De Deyn PP, Van Broeckhoven C, Livingston G, Bass NJ, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Al-Chalabi A, Shaw CE, Tsolaki M, Singleton AB, Guerreiro R, Mühleisen TW, Nöthen MM, Moebus S, Jöckel KH, Klopp N, Wichmann HE, Pankratz VS, Sando SB, Aasly JO, Barcikowska M, Wszolek ZK, Dickson DW, Graff-Radford NR, Petersen RC; Alzheimer's Disease Neuroimaging Initiative, van Duijn CM, Breteler MM, Ikram MA, DeStefano AL, Fitzpatrick AL, Lopez O, Launer LJ, Seshadri S; CHARGE consortium, Berr C, Campion D, Epelbaum J, Dartigues JF, Tzourio C, Alpérovitch A, Lathrop M; EADI1 consortium, Feulner TM, Friedrich P, Riehle C, Krawczak M, Schreiber S, Mayhaus M, Nicolhaus S, Wagenpfeil S, Steinberg S, Stefansson H, Stefansson K, Snaedal J,

Björnsson S, Jonsson PV, Chouraki V, Genier-Boley B, Hiltunen M, Soininen H, Combarros O, Zelenika D, Delepine M, Bullido MJ, Pasquier F, Mateo I, Frank-Garcia A, Porcellini E, Hanon O, Coto E, Alvarez V, Bosco P, Siciliano G, Mancuso M, Panza F, Solfrizzi V, Nacmias B, Sorbi S, Bossù P, Piccardi P, Arosio B, Annoni G, Seripa D, Pilotto A, Scarpini E, Galimberti D, Brice A, Hannequin D, Licastro F, Jones L, Holmans PA, Jonsson T, Riemenschneider M, Morgan K, Younkin SG, Owen MJ, O'Donovan M, Amouyel P, Williams J (2011). Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet. 43, 429-435.

Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, Pearson JV, Stephan DA, Nelson SF, Craig DW (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet. 4**,** e1000167.

Hu X, Pickering E, Liu YC, Hall S, Fournier H, Katz E, Dechairo B, John S, Van Eerdewegh P, Soares H (2011). Meta-analysis for genome-wide association study identifies multiple variants at the BIN1 locus associated with late-onset Alzheimer's disease. PLoS One. 6**,** e16616.

Huang DW, Sherman BT, Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols. 4**,** 44-57.

Huang R, Jaritz M, Guenzl P, Vlatkovic I, Sommer A, Tamir IM, Marks H, Klampfl T, Kralovics R, Stunnenberg HG, Barlow DP, Pauler FM (2011). An RNA-Seq strategy to detect the complete coding and non-coding transcriptome including full-length imprinted macro ncRNAs. PLoS One. 6**,** e27288.

Ikeda Y, Abe-Dohmae S, Munehira Y, Aoki R, Kawamoto S, Furuya A, Shitara K, Amachi T, Kioka N, Matsuo M, Yokohama S, Ueda K (2003). Posttranscriptional regulation of

human ABCA7 and its function for the apoA-I-dependent lipid release. Biochem Biophys Res Commun. 311, 313–318.

Iwamoto N, Abe-Dohmae S, Sato R, Yokoyama S (2006). ABCA7 expression is regulated by cellular cholesterol through the SREBP2 pathway and associated with phagocytosis. J Lipid Res. 47, 1915–1927.

Jehle AW, Gardai SJ, Li S, Linsel-Nitschke P, Marimoto K, Janssen WJ, Vandivier RW, Wang N, Greenberg S, Dale BM, Qin C, Henson PM, Tall AR (2006). ATP-binding cassette transporter A7 enhances phagocytosis of apoptotic cells and associated ERK signaling in macrophages. J Cell Biol. 174, 547–556.

Kanehisa M, Goto S (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28**,** 27-30.

Kapur K, Jiang H, Xing Y, Wong WH (2008). Cross-hybridization modeling on Affymetrix exon arrays. Bioinformatics. 24**,** 2887-2893.

Kellett KA, Hooper NM (2009). Prion protein and Alzheimer disease. Prion. 3**,** 190-194.

Kim WS, Guillemin GJ, Glaros EN, Lim CK, Garner B (2006). Quantitation of ATP-binding cassette subfamily-A transporter gene expression in primary human brain cells. Neuroreport. 17, 891–896.

Kim WS, Weickert CS, Garner B (2008). Role of ATP-binding cassette transporters in brain lipid transport and neurological disease. J Neurochem. 104:1145-1166.

Khatri P, Sirota M, Butte AJ (2012). Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol. 8**,** e1002375.

Kim KH, Moon M, Yu SB, Mook-Jung I, Kim JI (2012). RNA-Seq analysis of frontal cortex and cerebellum from 5XFAD mice at early stage of disease pathology. J Alzheimers Dis. 29**,** 793-808.

Konradi C, Eaton M, MacDonald ML, Walsh J, Benes FM, Heckers S (2004). Molecular evidence for mitochondrial dysfunction in bipolar disorder. Arch Gen Psychiatry. 61**,** 300-308.

Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B, Letenneur L, Bettens K, Berr C, Pasquier F, Fievet N, Barberger-Gateau P, Engelborghs S, De Deyn P, Mateo I, Franck A, Helisalmi S, Porcellini E, Hanon O, de Pancorbo MM, Lendon C, Dufouil C, Jaillard C, Leveillard T, Alvarez V, Bosco P, Mancuso M, Panza F, Nacmias B, Bossu P, Piccardi P, Annoni G, Seripa D, Galimberti D, Hannequin D, Licastro F, Soininen H, Ritchie K, Blanche H, Dartigues JF, Tzourio C, Gut I, Van Broeckhoven C, Alperovitch A, Lathrop M, Amouyel P (2009). Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet. 41**,** 1094-1099.

Lerch JK, Kuo F, Motti D, Morris R, Bixby JL, Lemmon VP (2012). Isoform diversity and regulation in peripheral and central neurons revealed through RNA-Seq. PLoS One. 7**,** e30417.

Li YY, Cui JG, Hill JM, Bhattacharjee S, Zhao Y, Lukiw WJ (2011). Increased expression of miRNA-146a in Alzheimer's disease transgenic mouse models. Neurosci Lett. 487**,** 94-98.

Lin M, Pedrosa E, Shah A, Hrabovsky A, Maqbool S, Zheng D, Lachman HM (2011). RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. PLoS One. 6**,** e23356.

Lobo A, Launer LJ, Fratiglioni L, Andersen K, Di Carlo A, Breteler MM, Copeland JR, Dartigues JF, Jagger C, Martinez-Lage J, Soininen H, Hofman A (2000). Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. Neurology. 54**,** S4-9.

Lu T, Pan Y, Kao SY, Li C, Kohane I, Chan J, Yankner BA (2004). Gene regulation and DNA damage in the ageing human brain. Nature. 429**,** 883-891.

Martin JA, Wang Z (2011). Next-generation transcriptome assembly. Nat Rev Genet. 12**,** 671-682.

Mills JD, Janitz M (2012). Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases. Neurobiol Aging. 33**,** 1012 e1011-1024.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 40**,** 1413-1415.

Pavlidis P, Qin J, Arango V, Mann JJ, Sibille E (2004). Using the gene ontology for microarray data mining: a comparison of methods and application to age effects in human prefrontal cortex. Neurochem Res. 29**,** 1213-1222.

Peng Z, Cheng Y, Tan BC, Kang L, Tian Z, Zhu Y, Zhang W, Liang Y, Hu X, Tan X, Guo J, Dong Z, Liang Y, Bao L, Wang J (2012). Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat Biotechnol. 30**,** 253-260.

Satoh J (2012). Molecular network of microRNA targets in Alzheimer's disease brains. Exp Neurol. 235**,** 436-446.

Shendure J (2008). The beginning of the end for microarrays? Nat Methods. 5**,** 585-587.

Stolzing A, Grune T (2004). Neuronal apoptotic bodies: phagocytosis and degradation by primary microglial cells. FASEB J. 18, 743–745.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 102**,** 15545-15550.

Sutherland GT, Janitz M, Kril JJ (2011). Understanding the pathogenesis of Alzheimer's disease: will RNA-Seq realize the promise of transcriptomics? J Neurochem. 116**,** 937-946.

Tollervey JR, Wang Z, Hortobagyi T, Witten JT, Zarnack K, Kayikci M, Clark TA, Schweitzer AC, Rot G, Curk T, Zupan B, Rogelj B, Shaw CE, Ule J (2011). Analysis of alternative splicing associated with aging and neurodegeneration in the human brain. Genome Res. 21**,** 1572-1582.

Toung JM, Morley M, Li M, Cheung VG (2011). RNA-sequence analysis of human B-cells. Genome Res. 21, 991-998.

Trapnell C, Pachter L, Salzberg SL (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 25**,** 1105-1111.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012). Differential gene and transcript expression analysis of RNA-Seq experiments with TopHat and Cufflinks. Nature Protocols. 7**,** 562-578.

Twine NA, Janitz K, Wilkins MR, Janitz M (2011). Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. PLoS One. 6**,** e16266.

Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature. 474**,** 380-384.

Wang S, Qaisar U, Yin X, Grammas P (2012). Gene expression profiling in Alzheimer's disease brain microvessels. J Alzheimers Dis. 31**,** 193-205.

Wang Z, Gerstein M, Snyder M (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 10**,** 57-63.

Williams C, Mehrian Shai R, Wu Y, Hsu YH, Sitzer T, Spann B, McCleary C, Mo Y, Miller CA (2009). Transcriptome analysis of synaptoneurosomes identifies neuroplasticity genes overexpressed in incipient Alzheimer's disease. PLoS One. 4**,** e4936.

Xing Y, Stoilov P, Kapur K, Han A, Jiang H, Shen S, Black DL, Wong WH (2008). MADS: a new and improved method for analysis of differential alternative splicing by exon-tiling microarrays. RNA. 14**,** 1470-1479.

Ziats MN, Rennert OM (2011). Expression profiling of autism candidate genes during human brain development implicates central immune signaling pathways. PLoS One. 6**,** e24691.