META ANALYSIS OF ALZHEIMER'S DISEASE AT THE GENE EXPRESSION LEVEL

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

HAMİT İZGİ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGY

FEBRUARY 2017

Aproval of the thesis:

META ANALYSIS OF ALZHEIMER'S DISEASE AT THE GENE EXPRESSION LEVEL

submitted by **HAMİT İZGİ** in partial fulfillment of the requirements for the degree of **Master of Science in Biology Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Orhan Adalı
Head of Department, **Biology** _____

Assoc. Prof. Dr. Mehmet Somel
Supervisor, **Biology Dept., METU** _____

Examining Committee Members:

Assoc. Prof. Dr. Özlen Konu
Dept. of Mol. Bio. and Gen., İhsan Doğramacı Bilkent Uni. _____

Assoc. Prof. Dr. Mehmet Somel
Biology Dept., METU _____

Assist. Prof. Dr. Nurcan Tunçbağ
Informatics Inst., METU _____

Assoc. Prof. Dr. Sreeparna Banerjee
Biology Dept., METU _____

Assoc. Prof. Dr. Çağdaş Devrim Son
Biology Dept., METU _____

Date: 03.02.2017

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : HAMİT İZGİ

Signature :

ABSTRACT

META ANALYSIS OF ALZHEIMER'S DISEASE AT THE GENE EXPRESSION LEVEL

İZGİ, HAMİT
M.S., Department of Biology
Supervisor : Assoc. Prof. Dr. Mehmet Somel

February 2017, 119 pages

In this study, publicly available microarray gene expression datasets are used to investigate common gene expression changes in different postmortem brain regions in Alzheimer's Disease (AD) patients compared to control subjects, and to find possible functional associations related to these changes. The hypothesis is that pathogenesis of the disease converges into common patterns of dysregulation/alteration or dysfunction in molecular pathways across different brain regions in AD. In total, I studied 13 datasets, one of which was excluded from the analysis in quality checks, resulting in 12 datasets spanning 7 different brain regions. Instead of using the standard approach to identify differentially expressed genes in each dataset independently, I used an alternative scheme, focusing on shared trends across all datasets, and testing their significance using cross-dataset structured permutations. Among more than 8000 common genes in all 12 datasets, I identified those showing shared upregulated (631) or downregulation (580) trends in AD across all datasets, which was highly significant compared to permutations. I then performed GO Biological Process enrichment analysis on both gene sets. There were 343 GO BP categories enriched for upregulated genes and 94 GO BP categories enriched for downregulated genes. Among 343 GO categories enriched for upregulated genes, the most noticeable ones include protein modification, differentiation, and the cell cycle. Furthermore, cell-cell signaling, synaptic activity and energy metabolism

related pathways are enriched in downregulated genes. These findings are in line with the effects of pathological changes in AD and suggests that different brain regions share common pathways deregulated by AD.

Keywords: Alzheimer's Disease, gene expression, microarray, brain

ÖZ

ALZHEIMER HASTALIĞININ GEN ANLATIMI DÜZEYİNDE META ANALİZİ

İZGİ, HAMİT
Yüksek Lisans, Biyoloji Bölümü
Tez Yöneticisi : Doç. Dr. Mehmet Somel

Şubat 2017, 119 sayfa

Bu çalışmada, yayınlanmış mikrodizin gen ifade veri setleri kullanılarak, kontrol gruplarına kıyasla Alzheimer hastalarının postmortem beyin bölgelerindeki ortak gen ifade değişiklikleri araştırılmış ve bu ortak değişimlerin muhtemel fonksiyonel sonuçları saptanmaya çalışılmıştır. Çalışmadaki hipotezimiz, Alzheimer hastalığı gelişimi sırasında farklı beyin bölgelerinde ortak bir şekilde moleküler yolaklarda dengenin bozulmasına, değişmesine ve işlevsel bozukluğa neden olmasıdır. Toplamda 13 veri seti incelenmiş, bir tanesinin kalite kontrolü sonucu çalışmadan çıkarılması suretiyle 7 farklı beyin bölgesini kapsayan 12 veri seti analiz edilmiştir. Çalışmada, ayrı ayrı her bir veri setinde farklı gen ifadesi gösteren genlerin tespiti şeklindeki standart yaklaşım yerine, veri setleri arasındaki ortak gen ifadesi değişimi eğilimlerinin tespiti ve bunların istatistiksel anlamının yapılandırılmış permütasyonlar yoluyla belirlenmesi yöntemi kullanılmıştır. Veri setleri arasında ortak 8000'in üzerindeki gen arasından Alzheimer'de anlatımı ortak biçimde artma eğilimi gösteren 631 tane ve anlatımı azalma eğilimi gösterene 580 tane gen tespit edilmiş ve bunların istatistiksel olarak yüksek derecede anlamlı olduğu belirlenmiştir. Daha sonra bu genler kullanılarak, GO Biyolojik İşlev zenginleştirme analizi yapılmış ve anlatımı artan genlerin 343 GO kategorisinde, anlatımı azalan genlerin ise 94 GO kategorisinde zenginleştiği bulunmuştur. 343 GO kategorisi arasında en dikkat çekici olanlar, protein modifikasyonu, farklılaşma ve hücre dön-

güsüdür. Aynı zamanda, hücre-hücre sinyali, sinaptik aktivite ve enerji metabolizmasıyla ilişkili yolakların ise anlatımı azalan genlerde zenginleştiği gösterilmiştir. Bu çalışma, Alzheimer hastalığındaki patolojik değişikliklerin etkileriyle moleküler değişikliklerin aynı doğrultuda olduğu göstermekte ve farklı beyin bölgelerinin aslında benzer şekilde hastalıktan etkilendiğine işaret etmektedir.

Anahtar Kelimeler: alzheimer hastalığı, gen anlatımı, mikrodizin, beyin

to my beloved family

ACKNOWLEDGEMENTS

I would like to thank to many people who helped and supported me through my MSc. First and foremost, I would like to express my gratitude to my supervisor Mehmet Somel for the continuous support of my MSc study. I am grateful for his patience, motivation and guidance.

I also want to thank to my all lab friends for their endless support. I must especially thank to Melike Dönertaş, Gözde Turan and Poorya Parvizi. They were always helpful whenever I had questions. I again want to thank faithfully to Melike Dönertaş for her great perspective and fruitful discussions not only as a lab mate but also as a sincere friend. I want to thank to Ekin Sağlıcan, İdil Yet and Ezgi Özkurt for their suggestions and grammar corrections to my thesis. Many thanks to Melike Dönertaş for providing me with Latex template. I also want to thank my fellows, Melike Dönertaş, Buse İşbilir, Betül Taşkoparan and Ekin Sağlıcan.

I want to thank Scientific and Technical Research Council of Turkey, TUBITAK, for financially supporting me for one year through 1002 project scholarship with "215Z053" code and "Sistem Biyolojisi Yaklaşımı İle Yaşlanan Beyin Dokusunda Hücresel Ağ Entropisinin Tanımlanması" title.

Lastly, I want to thank to my family for their love and support. This work would not be possible without their love and care for me.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

AD	Alzheimer's Disease
Aβ	Amyloid-β
NFT	Neurofibrillary Tangles
RMA	Robust Multi-Array Analysis
HC	Hippocampus
PFC	Pre-Frontal Cortex
TC	Temporal Cortex
SFG	Superior Frontal Gyrus
PC	Posterior Cingulate
CRB	Cerebellum
VCX	Visual Cortex
EC	Entorhinal Cortex

CHAPTER 1

INTRODUCTION

1.1 Alzheimer's Disease

Alzheimer's Disease (AD) is a complex and heterogeneous neurodegenerative disease that involves problems with memory, thinking and behavior, and is the most common form of dementia. Disease progression starts with mild symptoms and become severe in later stages. Prevalence of AD is constantly increasing in ageing populations worldwide (Barnes & Yaffe, 2011). Currently, there is no cure for AD but there are treatments that can reduce or slow down symptoms and improve quality of life. Meanwhile, an increasing amount of research on AD worldwide is being done to understand and find treatment for AD.

1.1.1 Discovery of Alzheimer's Disease

Alzheimer's Disease, or Alzheimer's, was discovered by German psychiatrist Alois Alzheimer, in 1906. He is the first in publishing the histological alterations in brain pathology of a 51-year-old woman who had symptoms of short-term memory loss and other strange behavioral symptoms (e.g. aggressiveness, crying and progressive confusion). After the patient's death, Alzheimer performed an autopsy to investigate the patient's brain and found histological alterations later described as plaques and neurofibrillary tangles. His findings did not get much attention in the German psychiatrists congress in 1906 but he did not give up his research. Three other cases were reported between 1906 and 1909 by Alzheimer and his colleague.

Kraepelin, a coworker of Alzheimer, introduced the term Alzheimer's Disease in new edition of her textbook in 1910 (Hippius & Neundorfer, 2003). Since that time, the term has been generally used. Because the disease was very rare, findings

of Alzheimer did not get much attention in the following years until 1970s. Later in 1976, Alzheimer's Disease was recognized as the most common cause of dementia by neurologist Robert Katzman (Katzman et al., 1976).

1.1.2 Signs and Symptoms

One of the biggest challenges for the diagnosis of AD is that damage in the brain starts years before the symptoms appear. In these early stages, people continue their everyday lives without any observable cognitive and memory problems. However, amyloid plaques and neurofibrillary tangles gradually accumulate in the brain. Initial damage starts to take place in hippocampus region of the brain, responsible for the formation of memories. As the disease progresses, other brain regions are affected too. Eventually, neurons become unable to function and lose connection with other neurons and die.

In the early stages, the first cognitive symptom that appears is memory problems, although this is variable from person to person. For the majority of people with AD, short-term memory loss starts interfering with daily life. Since hippocampus is the first affected brain region and it is responsible for day-to-day memory, long-term memories are unaffected at this stage. Memory loss increasingly interferes with daily life as the disease progresses. Planning and problem solving skills become challenging. Visual skills are also affected. Reading a page and judging distance gets harder. In moderate stages, memory loss and other cognitive difficulties become more severe. Daily tasks become more challenging to perform. The patients may have difficulties to recall family and friends. Speaking difficulties arise due to inability to recall memory. Finally, in the late stages of the disease, people become completely dependent on others. They cannot perform very simple tasks on their own. Language drops to simple phrases and words. Amyloid plaques and neurofibrillary tangles spread over the brain in this stage. Brain tissue shrinks significantly.

1.1.3 Pathophysiological Changes in AD

Although the whole brain is affected by AD at later stages, several regions are especially vulnerable starting from the very early stages. That is, disease progression

is not uniform throughout the brain. AD pathology starts in brain regions related to learning, memory and perception. Hippocampus, amygdala, entorhinal and cingulate cortices are the main regions initially affected by AD (Braak & Braak, 1991; Hampel et al., 2008; Loring, Wen, Lee, Seilhamer & Somogyi, 2001) As the disease progresses, several other brain regions start to display metabolic and pathological differences including frontal cortex, visual cortex, temporal cortex and cerebellum (Brewer & Barton, 2014; DeKosky & Scheff, 1990).

Neuronal and synaptic loss are the characteristics of Alzheimer's Disease in cortex and other certain regions of the brain. Pathological hallmarks of AD are amyloid plaques deposited in extracellular matrix and neurofibrillary tangle (NFT) formation in cell body. Amyloid plaques are formed by heterogeneous amyloid-β peptides, which are the products of proteolytic cleavage of amyloid precursor protein (APP). Through cleavage by secretases at different sites, APP yields two species of amyloid-β peptides, which are the major constituents of amyloid plaques; Aβ_{1-40} and Aβ_{1-42}. Aβ_{1-42}, more hydrophobic and amyloidogenic, is the major component and aggressive form of amyloid peptide (Selkoe, 1998). Mechanism of APP processing and formation of amyloid-β variants are discussed below (**See Section 1.2.2**).

The second hallmark of AD is the intracellular neurofibrillary tangle formation, which is not specific to AD. Several other neurodegenerative diseases, known as tauopathies, also involve NFT. The major component of NFT is the tau protein. Its physiological function is to stabilize and assemble microtubules (Weingarten, Lockwood, Hwo & Kirschner, 1975). Abnormal hyperphosphorylation and aggregation of tau protein leads to self-assembly and the formation of fibrillary tangles (Alonso 2001).This intracellular fibrillary structure prevents normal function of tau leading to impairment in microtubule integrity, which in turn results in neural dysfunction and cell death (Duckley 2006).

1.2 Causes of AD

The underlying mechanism of AD is not fully understood. The causes of AD are suggested to be combinations of genetic and environmental factors (Bird, 1993).

Genetically, Alzheimer's Disease is divided into three categories; chromosomal, familial and sporadic. Chromosomal type refers to only the Down Syndrome, or trisomy 21. Chromosome 21 contains the *APP* gene, which has essential role in Alzheimer's Disease (**Section 1.2.2**). Sporadic AD is the most common form of AD accounting for approximately 75% of the cases. It contains all the cases with non-familial AD, excluding chromosomal case. Familial AD is the form of the disease when at least two AD cases occur in a family. It can be either early-onset or late-onset, depending on the age of the patient when the disease appears (Bird, 1993).

Causes of Alzheimer Disease

Cause	% of Cases
Chromosomal (Down syndrome)	<1%
All familial	~25%
• Late-onset familial (AD2)	15%-25%
• Early-onset familial AD (AD1, AD3, AD4)	<2%
Unknown (includes genetic/environment interactions)	~75%

Figure 1.1: Percentages of AD according to GeneReviews in NCBI Bookshelf, updated in 2015 (Bird, 1993)

1.2.1 Chromosomal Cause

Down syndrome (DS) is a genetic disorder caused by trisomy of chromosome 21. Chromosome 21 carries the *APP* gene which plays a vital role in AD neuropathology. Therefore, the vast majority of people with DS, being trisomic for *APP*, develop AD pathology after 40 years of age (Bird, 1993). It has been shown that APP over-expression might lead to accumulation of amyloid-β in the brains of children with DS (Leverenz & Raskind, 1998). In addition, there can be other factors contributing to development of AD pathology in DS. It is suggested that other genes in chromosome 21 might also play role. The trisomy of oxidative stress-related gene

located in chromosome 21, superoxide dismutase, may be related to dysfunction of oxidative damage repair in DS and thus might contribute to development of AD pathology (Lott & Head, 2001).

The association between AD and DS is assumed to be the lifelong over-expression of *APP*, which results in overproduction of amyloid-β protein in brains of DS patients who are trisomic for *APP* gene. This hypothesis is supported by a study which states that no neuropathological evidence of AD was observed in a 78 year-old woman who carried a partial copy of chromosome 21, not including the *APP* gene (Prasher et al., 1998).

1.2.2 Familial AD

Familial AD (FAD) refers to the cases where at least two people in a family have been diagnosed with AD. About 25% of AD cases are familial, which are further divided into early-onset and late-onset types. If several members of a family are diagnosed as AD before mean age 65, it is referred as early-onset familial AD (EOFAD). But the age of onset threshold is a somewhat arbitrary decision. Some studies use 60 years or 70 years to diagnose EOFAD. However, age of onset in many cases classified as EOFAD is younger than 60 years. Approximately, 60% of early-onset AD cases, which accounts for 1%-6% of all AD, are familial (Campion et al., 1999). Mutations of three genes are well documented for EOFAD (see below). Late-onset familial AD (LOFAD) is more common, accounting for approximately 25% of all AD cases. LOFAD is a complex disease involving several susceptible genes. The *APOE*-ϵ4 allele is a well characterized gene associated with LOFAD.

Early-Onset Familial AD

Early-onset familial AD (EOFAD) refers to AD cases where age of onset is before 65 years through generations in a family. It is inherited in an autosomal dominant manner. Mutations in three genes, *APP, PSEN1* and *PSEN2*, are associated with early-onset familial AD. More than 30 *APP* mutations, 179 *PSEN1* and 14 *PSEN2* mutations have been discovered in early-onset, autosomal dominant AD (O'Brien

& Wong, 2010). However, other autosomal dominant early-onset AD cases, in which those three genes do not carry mutations, have also been reported. Thus, there might be mutations in additional genes that play role in AD.

APP

APP, the amyloid precursor protein, is a large gene located on the long arm of chromosome 21. It spans about 240kb and contains 18 exons (Yoshikai, Sasaki, Dohura, Furuya & Sakaki, 1990). It is a highly conserved gene, coding for a single pass transmembrane protein with a large extracellular domain (Tharp & Sarkar, 2013) (**Figure 1.2a**). It is expressed in many tissues and concentrated on synapses of neurons. *APP* has several alternative splicing isoforms ranging from 365 to 770 aminoacids and some of them are preferentially expressed in neurons. The reason and functional importance behind this tissue-specific alternative splicing of *APP* is not well understood.

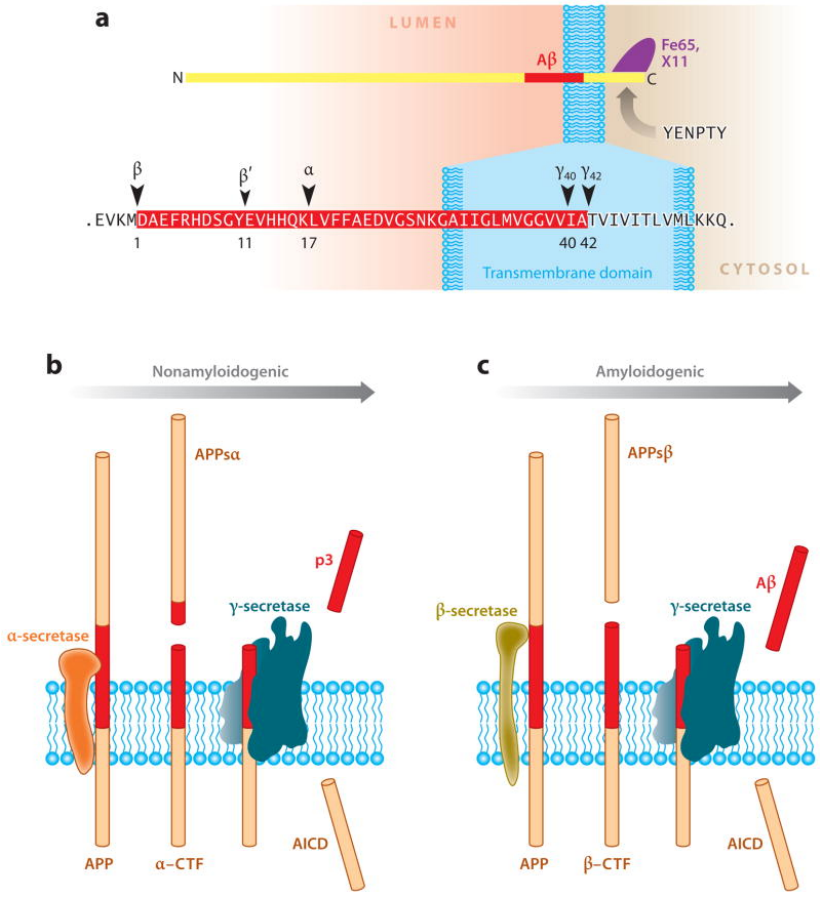

Figure 1.2: Structure of APP (a) and the Aβ production process (c) (Amyloid Precursor Protein Processing and Alzheimer's Disease (O'Brien & Wong, 2010).

APP is subject to extensive post-translational modifications including glycosylation, phosphorylation and proteolytic processing. Full length APP is proteolytically cleaved by three proteinase enzymes called α-, β- and γ- secretases via two pathways. In nonamyloidogenic pathway, APP is first cut by α-secretase inside the amyloid-β (Aβ) peptide region. This cleavage results in two fragments, one of which is released as large secreted extracellular domain (sAPP-α) (**Figure 1.2b**). The other fragment is cut by γ-secretase to release two peptides, which are degraded rapidly. sAPP production through α-secretase cleavage is a constitutive process (Esch et al., 1990). ADAM10 (a disintegrin and metalloproteinase) is one of the three predicted α-secretases (Lammich et al., 1999). Overexpression of ADAM10 increases the production of sAPP protein while decreasing Aβ production and plaque formation. Likewise, production of mutant ADAM10 proteins can

increase Aβ pathology (Lammich et al., 1999). One of the pathological hallmarks of AD is the accumulation of the Aβ peptide. Aβ is produced via the amyloidogenic pathway (**Figure 1.2c**). Cleavage of APP by sequential β- and γ-secretases yields a mixture of Aβ peptides with different lengths. There are two main Aβ species AB_{1-40} (%90) and AB_{1-42}(%10), the latter being more aggregation-prone and predominantly present in amyloid plaques in brains of AD patients (Van Cauwenberghe, Van Broeckhoven & Sleegers, 2016). Majority of the mutations of *APP* gene favors production of aggressive-prone AB_{1-42} peptide. Nonetheless, there is no clear association between cognitive decline and amyloid plaque load. A study reports that amyloid plaque deposition does not correlate with cognitive impairment (Giannakopoulos et al., 2003).

Although some of them are speculative, several functions are attributed to APP. It has been shown that the extracellular domain of APP binds to a neuronally secreted glycoprotein, F-spondin and regulates amyloid-β production and downstream signaling (Ho & Südhof, 2004). More evidence exists for the function of APP as a cell adhesion protein. In vivo studies in mouse show homo- and heterodimerization of APP family proteins, suggesting a role in trans-cellular adhesion (Soba et al., 2005). In addition, colocalization of APP and integrins in rat primary culture neurons supports this possible function (Yamazaki, Koo & Selkoe, 1997). Another important function attributed to APP is its role in neurite growth and synaptogenesis. In cell culture studies, expression of APP is shown to be upregulated in neuronal differentiation (Hung, Koo, Haass & Selkoe, 1992). Furthermore, upregulation of APP after traumatic brain injury has been shown in mammalian and Drosophila brain (Leyssen et al., 2005; Van den Heuvel et al., 1999) suggesting a repair role. In line with these observations, other studies have shown that APP plays an important role in neuron viability and synaptic activity (Hérard et al., 2006; Perez, Zheng, Van der Ploeg & Koo, 1997). The important roles assigned to the intracellular domain of APP are notably axonal transport, cell signaling and synapse remodeling. Changes in the ratio of isoforms expressed in neurons as well as the phosphorylation sites of APP protein are associated with Alzheimer's Disease (Matsui et al., 2007). APP can be phosphorylated at multiple sites. Specifically, Thr668 phosphorylated APP is shown to increase in AD compared to control subjects suggesting that APP phosphorylation may regulate Aβ production and eventually contribute to AD pathogenesis (Lee et al., 2003).

PSEN1 and PSEN2

Since the discovery of mutations in *PSEN* genes in early onset familial AD (Sherrington et al., 1995), many studies have focused on FAD-linked mutations and biology of these genes. *PSEN1* and *PSEN2*, highly homologous genes, code for presenilin-1 and presenilin-2 proteins in humans, respectively. The *PSEN1* gene is located on chromosome 14 and *PSEN2* is located on chromosome 1. While *PSEN2* mutations are rare, *PSEN1* mutations are very frequent in autosomal dominant AD.

Both presenilin-1 and presenilin-2 proteins are components of the γ-secretase complex, which is a multi-subunit, integral membrane protein (Vetrivel, Zhang, Xu & Thinakaran, 2006). γ-secretase catalyzes the cleavage of single-pass membrane proteins, including Notch and APP. As mentioned above (in the APP section), γ-secretase plays a role both in non-amyloidogenic and amyloidogenic pathways of APP. In amyloidogenic pathway, β-secretase produces a short APP fragment, which is further cleaved by γ-secretase to produce Aβ peptides in different lengths.

Mutations in *PSEN1* and *PSEN2* genes alter the proteolytic activity of γ-secretase resulting in an increased ratio of Aβ_{1-42} to Aβ_{1-40}. *PSEN1* mutations are responsible for the most severe forms of familial AD. All these mutations associated with AD are autosomal dominant, which means mutation in one allele is sufficient to shift development of AD to earlier ages. Mechanistically, these mutations do not interfere with assembly of γ-secretase complex, which means it is proteolytically active. However, complete loss of function mutations in *PSEN1* and *PSEN2* genes cause severe disorders but have no relation to neurodegeneration or AD (Wang et al., 2010). Although mutations in the *APP* gene have role in Aβ pathogenesis, proteolytic cleavage of APP by γ-secretase (which includes presenilin proteins in its multi-subunit structure) is the most important step affecting Aβ deposition such that mutations in *PSEN* genes alters γ-secretase activity by increasing the ratio of Aβ_{1-42} to Aβ_{1-40} (Cruts & Van Broeckhoven, 1998).

To sum up, early-onset familial AD is associated with mutations in three genes, *APP*, *PSE1* and *PSE2*. Mutations in the *PSEN1* gene are the most common cause,

followed by those in the *APP* gene. Mutations in the *PSEN2* gene, associated with early-onset familial AD, are rare. However, there are also rare cases of EOFAD that cannot be associated with one of these three genes. Thus, there might be mutations in other genes involved in EOFAD.

Late-Onset Familial AD

Familial AD accounts for approximately 25% of all AD cases and more than 75% of them are late-onset. Late-onset AD is a more complex form of AD that is multifactorial and may involve multiple genes. There is not a gene identified where dominant mutations cause late-onset familial AD, as in the case in early-onset familial AD. Still, the *APOE*-ϵ4 allele is the well-described and most studied risk factor for late-onset AD. Nonetheless, *APOE*-ϵ4 allele is also considered to be risk factor for early-onset AD shifting age of onset to early ages (Khachaturian, Corcoran, Mayer, Zandi & Breitner, 2004). Although early-onset and late-onset forms of AD are defined separately, the only distinction between them is the age of onset. However, this distinction is vague in real life because AD is a complex disease, which develops through a continuous accumulation of pathological and physiological processes with the patient's age. Moreover, a study reports that about 25% of families with late-onset AD also have a relative with early-onset AD (Brickell et al., 2006). Therefore, AD is accepted as a one single disease apart from the differences in genetic cause and age of onset.

The *APOE*-ϵ4 allele is associated with late-onset familial and sporadic forms of AD. Although, the link between positive family history and presence of *APOE*-ϵ4 allele is very high, there was no evidence of transmission of AD though *APOE* gene within families (Jarvik, Larson, Goddard, Schellenberg & Wijsman, 1996). The role of *APOE* allelic variants and other genes susceptible in late-onset AD are discussed below.

1.2.3 Sporadic AD

Individuals with AD who have no family history are called sporadic cases. Sporadic AD can occur both in early and late ages and accounts for majority of all AD cases.

It is a complex and heterogeneous disease suggested to be the result of combinatorial effects of ageing, genetic and environmental factors. While familial AD is frequently characterized by mutations in *APP*, *PSEN1* or *PSEN2*, there is no exact pathogenesis of sporadic AD. Although the genetic backgrounds of familial and sporadic AD are different, they are clinically indistinguishable. *APOE*-ϵ4 allele is the main characterized risk factor for AD. However, more than 40% of AD cases are reported not to carry the *APOE*-ϵ4 allele (Mayeux et al., 1998). The presence of *APOE*-ϵ4 allele is neither necessary, nor sufficient for the diagnosis of late-onset AD.

Apolipoprotein E

The *Apolipoprotein E* (*APOE*) gene is located on chromosome 19 and consists of four exons. It encodes for a 299 aminoacid long protein, which is a component of lipoproteins. ApoE is an essential protein in lipid homeostasis and acts by mediating lipid and cholesterol transport. It is expressed in many tissues, with highest expression in the liver followed by the brain. Upon neuronal degeneration, apoE collects remaining lipids and redistributes them to cells in need of lipids for membrane repair or myelination of new axons (Huang, 2006). The main cell type that expresses apoE in the brain is astrocyte, and to some extent microglia (Grehan, Tse & Taylor, 2001). Neurons in transgenic mouse models also express human apoE at low levels in response to excitotoxic injury (Xu et al., 2006).

APOE is a polymorphic gene containing four different isoforms (or alleles), three of them common. These three, *APOE*-ϵ2, *APOE*-ϵ3 and *APOE*-ϵ4 differ from each other by one aminoacid residue. Different isoforms display specific activities. *APOE*-ϵ3 and *APOE*-ϵ2 are suggested to be effective in maintenance and repair of neuronal cells while *APOE*-ϵ4 may have an opposite effect (Mahley, Weisgraber & Huang, 2006). *APOE*-ϵ3 secreted from astrocytes also stimulates neurite growth and extension in mice hippocampus but *APOE*-ϵ4 has no such effect (Sun et al., 1998).

The *APOE*-ϵ4 allele is the major risk factor for both familial and sporadic late-onset AD but it is not sufficient alone to cause AD (Saunders et al., 1993). Gene

dose of *APOE*-ϵ4 allele has also important effect on age of onset. While the risk for AD is estimated to be three-fold higher for heterozygous carriers (*APOE*-ϵ3/ϵ4) than non-carriers, it is fifteen-fold for homozygous carriers (*APOE*-ϵ4/ϵ4) (Corder et al., 1993). Unlike ϵ4, ϵ2 allele has a protective effect against AD and delays age of onset. Risk of AD is estimated to be lowest in individuals bearing ϵ2/ϵ3 alleles (Corder et al., 1994).

Amyloid-β plaque deposition is one of the two main pathological hallmarks of AD. Its relation with apoE protein is well-studied. ApoE protein binds to amyloid-β peptide in an isoform-specific manner. ApoE2, apoE3 and apoE4 can bind to amyloid-β peptide to form stable complexes. However, apoE4 is shown to bind to amyloid-β more rapidly and aggressively, while apoE2 shows the least binding (Ma, Yee, Brewer, Das & Potter, 1994). This is consistent with the observation that increased amyloid plaque deposition in individuals with *APOE*-ϵ4 genotype.

Genome Wide Association Studies of Alzheimer's Disease

To find risk factors other than the *APOE*-ϵ4 genotype, Genome Wide Association Studies (GWAS) of AD have been conducted by multiple laboratories. A GWAS done on more than 2,000 AD and control subjects found two loci associated with AD risk: *APOJ* (encoding for apolipoprotein J) and *CR1* (encoding for a complement component), which may have role in clearance of amyloid-β peptide (Lambert et al., 2009). Two other GWAS also report *APOJ* and *CR1* loci as risk factors for AD, together with novel loci they identified, including *PICALM* and *BIN1* (Harold et al., 2009; Seshadri et al., 2010). However, none of the loci found in GWAS have a risk effect that can be comparable to that of *APOE*-ϵ4. Still, these studies reveal important findings on pathophysiological pathways related to AD. For instance, genes found in GWAS were clustered in three main pathways; cholesterol and lipid metabolism, inflammatory response and endosomal vesicle cycling (Van Cauwenberghe et al., 2016).

1.3 Transcriptome Change in Brain with AD

Transcriptome studies in AD have been drawing growing attention for nearly two decades. Most of these studies are aimed at discovering pathogenesis of the disease, biomarkers and cellular pathways correlated to AD progression by focusing on affected brain regions individually. One of the important limitations of these transcriptome studies is that technical artifacts, as well as cellular heterogeneity and other sources of uncontrolled biological or environmental variability between samples can influence the results. Despite such confounding and stochastic factors, their differences in the selection of tissue type or brain region, and frequently having small sample sizes, each study has provided significant results in their own context.

One of the earliest expression profiling studies was done in 2000 by Ginsberg and colleagues on AD patients bearing NFT in their CA1 neurons of hippocampus region. They compared relative expression of mRNAs in control and AD samples for $\sim$18,000 expressed sequence tags. They reported that compared to control, hippocampal neurons of AD patients have significantly reduced relative mRNA levels, some of which are implicated in AD pathology including synaptic proteins and phosphatases/kinases (Ginsberg, Hemby, Lee, Eberwine & Trojanowski, 2000).

Another transcriptome study was conducted on hippocampal CA1 tissue using 6 AD and 6 control samples (Colangelo et al., 2002). Authors reported that they analyzed 12,633 genes in two groups and found functional enrichment for both downregulated and for upregulated genes. They showed that among downregulated genes, there were signaling elements involved in synaptic plasticity, as reported before, and transcription factors. They also claimed that apoptotic and neuroinflammatory genes were activated in hippocampal CA1 neurons in the AD brain.

In another study published in 2009, the authors focused on the temporal cortex, which is among affected brains region in AD (Tan et al., 2010). They investigated temporal cortex expression profiles from 25 AD and 16 control subjects. They reported more than 5000 genes differentially expressed in AD. Functional analysis of these genes further supported the previous findings, as the authors showed enrichment in functional groups associated with AD including synaptic function, neuro-

transmission and neuroinflammation.

To investigate the relationship between severity of AD and gene expression change, a study conducted on 9 control and 22 AD subjects at different stages with AD (Blalock et al., 2004). They classified AD cases as incipient, moderate and severe using the MiniMental Status Examination (MMSE) and NFT scores. They showed that thousands of genes were correlated with MMSE and NFT measures, while a relatively small proportion of these genes were correlated with only control and incipient cases. Among genes correlated with incipient AD, they performed functional analysis and identified upregulation of transcription factors and signaling genes involved in proliferation, differentiation and tumor suppressors, as well as apoptosis and inflammation. This interesting finding supports the report of another previously published study which showed the activation of mitogenic signaling molecules in neurons of AD patients (Arendt et al., 2000). The latter study hypothesized that deregulation and/or activation of mitogenic molecules might lead neurons to attempt to re-enter cell cycle and dedifferentiate which eventually results in cell death.

Glucose metabolism in cerebral cortex is also associated with AD. Progressive decrease in metabolic rate of glucose is well-established in AD patients using regional positron emission tomography (PET) (Alexander, Chen, Pietrini, Rapoport & Reiman, 2002). The motivation that metabolic change in cortex might start before the onset of AD pathology and appearance of clinical features has led a group of scientists to investigate metabolically affected brain regions in AD patients, which could provide new insights into the pathogenesis of AD (Liang et al., 2008). Using normal and AD cases, they analyzed 6 brain regions vulnerable to metabolic change in glucose including cingulate cortex, middle temporal gyrus, hippocampus, entorhinal cortex, visual cortex and frontal cortex. They showed that AD cases had significantly decreased expression in 70% of nuclear genes encoding mitochondrial electron transport chain subunits in cingulate cortex, followed by 65% and 61% of those in middle temporal gyrus and hippocampus, respectively. They concluded that nuclear genes encoding mitochondrial energy metabolism are downregulated in neurons of AD cases, particularly in the cingulate cortex region.

Insulin is an important regulatory hormone in glucose metabolism. The fact that glucose metabolism is associated with AD raises the possibility that diabetes mel-

litus (DB) might also be risk factor for AD. A transcriptome study, published in 2014, investigated the association of AD-related gene expression changes with DB-related changes (Hokama et al., 2014). In this study, the authors analyzed microarray data from frontal cortex, temporal cortex and hippocampus in AD and control samples. They showed that hippocampus has the most significant gene expression alteration in AD patients. Also, they reported that expression of genes involved in DB and obesity also changed significantly in AD cases. They concluded that the decreased insulin signaling in the brain is a result of AD pathology.

Another study investigated the regional vulnerability in AD, specifically the CA1 and CA3 regions of hippocampus (Miller, Woltjer, Goodenbour & Horvath, 2013). CA1 neurons are severely affected in early stages of AD while CA3 neurons are less affected although they are structurally similar. The authors performed differential expression and co-expression analyses using the weighted gene co-expression network approach. They reported consistent results with previous findings and showed a link between disease status and brain region. Consistent with the observed pathological and region specific vulnerability, they found that the CA3 region has less abnormal expression compared to CA1. They also showed that genes downregulated in AD progression tend to be enriched in the CA3, region suggesting a link between transcriptome profile and a brain region's vulnerability to disease.

An interesting study investigating sporadic and familial early-onset AD was published by Antonell and colleagues in 2013. They investigated the expression profile of posterior cingulate brain region in early-onset sporadic and early-onset familial AD caused by *PSEN1* mutations using 7 patients for each type and 7 control subjects (Antonell et al., 2013). They reported 3183 and 3350 differentially expressed genes in these two types of AD, respectively, out of which 1916 genes were common. Interestingly, they did not find any differentially expressed gene between sporadic and familial cases. Performing functional analysis on differentially genes, they reported similar functional groups enriched in both AD cases including intracellular signaling pathways, axon guidance and synaptic plasticity. They suggested that although etiologies of these two groups of early-onset AD are different, the underlying mechanisms with different pathways might converge in a common final stage of the disease.

Network-based approaches to investigate molecular changes at transcriptome level could provide additional information about molecular interactions of complex disease nature. Two widescale independent studies, published in 2013 and 2014, respectively, employed gene regulatory networks to characterize molecular changes associated with late-onset AD (Narayanan et al., 2014; Zhang et al., 2013). In the former study, the authors showed that specific network structures were remodeled in AD and identified the key regulators of these networks. They reported *TYROBP* in immune/microglia module as a key regulator, which is thought to be involved in amyloid-b turnover and neuronal damage. The latter study focused on global alterations in co-regulation of genes in both AD and Huntington's Disease. They reported that networks of differentially co-expressed gene pairs showed increased or decreased correlation in AD compared to control. They claimed that increased correlation in these networks were more dominant than decreased correlations. They also identified a subnetwork enriched in chromatin organization and neural differentiation. Such network-based studies in the future may reveal further insights into molecular mechanisms in AD.

Almost all of the transcriptome studies focuses on a specific brain region. Integrating multiple datasets in a study to identify mutual and/or distinctive patterns among different brain regions might provide new insights about AD. A meta analysis of transcriptome study conducted on only hippocampus region investigated gene expression signatures in AD (Wruck, Schröter & Adjaye, 2016). The authors used hippocampus biopsies and iPSC-derived neurons using hierarchical clustering analysis. They showed over-representation of response to stress, regulation of cellular metabolic process and reactive oxygen species as well as two gene regulatory networks, FOXA1 and FOXA2, in the etiology of AD. However, the scope of this study was restricted to one brain region. Another meta analysis study was conducted by Puthiyedth N. and colleagues in 2016 using 6 different brain regions with AD (Puthiyedth, Riveros, Berretta & Moscato, 2016). They investigated the differentially expressed genes in each brain region and also identified common genes related to AD across all the regions. They found AD-related genes consistent with existing studies and also new candidate genes not previously related to AD. On the other hand, this study was performed using 6 datasets of a published transcriptome study (Liang et al., 2008), representing 6 different brain regions which share several samples from same individuals. Also, the total number of samples was not large

(44 samples). Thus, the common patterns across different brain regions might not have been reflected thoroughly in the study.

1.4 Research Objectives

As summarized above, there are several transcriptome studies on specific brain regions to elucidate the mechanisms underlying AD pathology, focusing on altered/ deregulated molecular pathways. Although each study provides valuable information about disease mechanisms in specific brain regions, none of them focuses on global changes across brain regions with AD. Thus, it remains unclear how much the same set of genes may be affected in their expression across the brain. A scan for shared gene expression patterns across different studies will thus be biologically interesting. Meanwhile, published transcriptome studies may also differ with respect to technical factors, such as the microarray platform type used, and may be affected by sampling error due to small sample sizes. Therefore, using a meta-analysis approach to integrate these datasets and to study common gene expression changes in AD might be a useful approach to reduce the effects of confounding factors and technical noise. As I mentioned above (**See Section 1.3**), previous meta-analysis studies were either restricted to only one brain region or their results might have been affected by small small size.

Here I used publicly available microarray gene expression datasets to investigate common gene expression changes in different postmortem brain regions in AD patients compared to control subjects, and to find possible functional associations related to these changes. Although the vulnerability levels of brain regions to AD is different and some regions are especially affected by AD, pathological and clinical findings, amyloid plaques and neurofibrillary tangles, are similar between them. I hypothesized that genetic risk factors, increased susceptibility with age and other unknown pathogenesis of the disease converge into a common dysregulation/alteration or dysfunction in molecular pathways among different brain regions affected by AD.

CHAPTER 2

MATERIAL AND METHOD

2.1 Datasets

This study was conducted using 9 different datasets published in the NCBI Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSExxx, GEO accession numbers are given in **Table 2.1**. Throughout my thesis, I will refer to the datasets after the first author and the publication year (**Table 2.1**). All gene expression datasets were microarray experiments produced using 3 major platform types: Affymetrix, Illumina Beadchip and Rosetta/Merck. Five Affymetrix datasets, Liang2007, Blalock2004, Hisayama2014, Tan2012 and Antonell2015, were analysed starting from the raw data (.CEL) files, which is possible using the free R packages "affy" (Gautier, Cope, Bolstad & Irizarry, 2004) or "oligo" (Carvalho & Irizarry, 2010). The four other datasets, Narayanan2014, Durrenberger2011, Miller2013 and Zhang2013, were not Affymetrix-based and the analysis of raw data from these experiments require commercial software; these were therefore analysed using from the processed "series matrix files" directly downloaded from GEO. In total, there were 856 samples spanning 6 brain regions (excluding the Durrenberger2011 dataset; (**Table 2.1**). The age distributions of control and AD-patients were visually compared to ensure they roughly match in each dataset, in order to avoid confounding between age-related and AD-related changes. To avoid such possible confounding, one young sample (22 years old) from Narayanan2014 and two young samples (22 and 25 years old) from Zhang2013 datasets were removed from analysis. Remaining samples in all datasets were above mean age 70 except Antonell2015, which had mean age 56.

Analysis of all datasets was performed in the R programming environment.

Table 2.1: Dataset column shows the first author of the article and published year that I have taken the dataset from. I will refer to each dataset using these names instead of GEO numbers. AD, Alzheimer's Disease; ND, No-Dementia. SFG; superior frontal gyrus, HC; hippocampus, PFC; prefrontal cortex, TC; temporal cortex, PC; posterior cingulate, EC; entorhinal cortex.

GEO accesion	**Dataset**	**Platform**	**Brain region**	**Total Sample**	**# of AD & ND**
GSE5281	Liang2007	HG-U133_plus_2	SFG	32	23AD-9ND
GSE1297	Belalock2004	HG-U133A	HC	29	20AD-9ND
GSE36980	Hisayama2014	HuGene-1_0-st	HC PFC TC	39	7AD-10ND 15AD-18ND 10AD-19ND
GSE37263	Tan2012	HuEx-1_0-st	TC	16	8AD-8ND
GSE39420	Antonell2015	HuGene-1_1-st	PC	14	7AD-7ND
GSE29378	Miller2013	Illumina humanHT-12V3.0 beadchip	HC	32	15AD-17ND
GSE26927	Durrenberger2011	Illumina humanHT-12V2.0 beadchip	EC	18	11AD-7ND
GSE33000	Narayanan2014	Rosetta/Merck Human44k 1.1	PFC	466	310AD-156ND
GSE44772	Zhang2013	Rosetta/Merck Human44k 1.1	PFC VCX CRB	228	129AD-99ND 129AD-99ND 129AD-99ND

2.2 Preprocessing of Gene Expression Datasets

Preprocessing refers to converting the raw signal data from a microarray's scanned image into quantitative estimates of gene expression level. Affymetrix microarrays are the most widely used platforms for gene expression studies. There are freely available R packages (also called libraries) for preprocessing of these arrays. For the datasets produced using Affymetrix microarray platforms, I used the raw data (.CEL) files, which contain light intensity information of each probe for each sample. I conducted the preprocessing using either the "oligo" or the "affy" package, according to the specific platform.

The Liang2007 and Blalock2004 datasets were preprocessed with the "affy" library. CEL files were loaded to R with the "ReadAffy" function to create an "affy" R object. Hokama2014, Tan2012 and Antonell2015 datasets were preprocessed with the "oligo" library. The "read.celfiles" function was used to read CEL files. Then the robust multi-array average (RMA) correction was performed using "rma" function in the "affy" or "oligo" libraries (**See 2.2.1**).

The four other datasets, Miller2013, Durrenberger2011, Narayanan2014 and Zhang 2013, are based on the Rosetta/Merck and Illumina beadchip arrays. These are less commonly used and there are no free R packages to conduct their preprocessing steps. Therefore, I used the "series matrix files" published in GEO for these datasets, which were already preprocessed by the authors. I continued analysis of all these 9 datasets with quantile normalization using "preprocessCore" package (**See 2.2.4**).

2.2.1 RMA

RMA (Robust Multi-Array Average) is a correction method for microarray experiments. There is a function in both "oligo" and "affy" R libraries to perform this correction. The "rma" function includes four steps, background correction, log transformation, normalization and summarization. Background correction removes noise and local artefacts from chip scan data, so that light intensities of probes are not affected by neighboring probe measurements (Gautier et al., 2004). Then, the function takes log2 values of light intensities and performs quantile normalization

(**See 2.2.4**). In the summarization step, intensity levels of multiple probes are combined into one probeset intensity, which represent gene expression levels for each transcript defined by the platform. The output of the "rma" function is an ExpressionSet object. Using the "exprs" function, ExpressionSet object is converted to a matrix where columns represent each sample and rows are probeset IDs.

R libraries "affy" and "oligo" were used for RMA correction of "HG-U133_plus _2" and "HG-U133A" platforms, and "HuGene-1_0-st", "HuEx-1_0-st" gene and "HuGene-1_1-st" platforms, respectively, as recommended by the package authors.

2.2.2 Probeset to Gene Conversion

In order to compare different datasets and perform functional analysis, probeset expression intensities should be converted (summarized) into a single gene expression intensity; thus a probeset should be mapped to a single gene ID. In many platforms, probeset to gene conversion is not always one-to-one. A probeset might correspond to more than one gene and multiple probesets might correspond to one gene. Here, probesets corresponding to more than one gene were removed from analysis (**Table 2.2**) since having multiple genes with the same expression value (represented by only one probeset value) would create a pseudoreplication problem. For multiple probesets corresponding to the same gene, for each sample, I took the mean expression level of those probesets to represent the corresponding gene. Alternatively, one of the multiple probesets could also be chosen to represent a gene; for instance, one could choose the probeset with the maximum average expression level across individuals in each dataset. The problem with this approach would be that the maximum to represent the genes individuals in a datasetr.e ID Bioconductorn. ama irsinsoftwareprobeset value chosen in one platform might not be present in other platforms at all. The probeset chosen might also differ from dataset to dataset. Therefore, choosing one probeset could potentially increase inconsistency between platforms and/or datasets.

For the Liang2007 and Blalock2004 datasets, biomaRt (Durinck et al., 2005) gene annotation data "hgu133plus2.db" and "hgu133a.db" (Carlson M, n.d.) were used respectively, while for the rest of the datasets biomaRt annotation data were not

available. Therefore, I used the platform-specific annotation file (called a GPL file) of each platform deposited in the GEO database (Edgar, Domrachev & Lash, 2002). GPL files of Narayanan2014, Durrenberger2011, Miller2013 and Zhang2013 datasets contain only Entrez Gene annotations. For those datasets, probeset IDs were first converted to Entrez Gene IDs, and then using biomaRt (version 84), Entrez Gene IDs were converted to Ensembl Gene IDs. Number of probesets and Ensembl Gene IDs in different platforms are summarized in **Table 2.2**.

It can be expected that for the same datasets, each subdataset (each brain region from the same publication) should have the same number of genes, since they are processed using the same platform. However, the "series matrix files" for Zhang2013, Narayanan2014, Miller 2013 and Durrenberger2011 datasets, which were downloaded from GEO as preprocessed by the authors, contained missing values. These genes I removed from each dataset, leading to different numbers of genes in the three Zhang2013 subdatasets (**Table 2.2 and 3.1**).

2.2.3 Log2 Transformation

Logarithmic conversion is widely used in transcriptome data analysis. The reasons are as follows: The bulk of light intensities obtained from microarray experiments are very low values, but there also exist fewer measurements of very high magnitudes. In addition, there is usually a strong linear relationship between mean and variance per gene. It is not convenient to perform analysis on such data, including visualization, or differential expression analysis using parametric models, such as ANOVA, which assume normality and equal variances among groups. Transforming all data to base-log2 brings values to similar orders of magnitude. It also removes the dependence between mean and variance. Usually, in datasets where there are values between 0 and 1, the value 1 is added to all data and then log2 transformation is done, to avoid minus values.

2.2.4 Quantile Normalization

Normalization has become a standard and essential step in preprocessing of microarray data. Technical variation is introduced during experimental steps and can-

not be avoided completely. Normalization techniques, such as quantile normalization, reduce technical and unwanted biological variation among samples that might lead to misinterpretation and false positive results. Quantile normalization assumes that expression levels across genes in all biological samples must have the same distribution, and global differences among samples (e.g. systematically higher expression levels or higher variance across the transcriptome) arise due to technical variations.

The principle of quantile normalization is as follows: gene expression levels in each sample are ranked (to be used in last step) and the original expression level data are sorted from lowest to highest. In the sorted data, for each quantile (gene or probeset), the corresponding expression levels across samples is taken, averaged, and the original values are substituted by this average. This ensures that all samples have the same distribution of expression values (but not the same expression level for each gene).

Quantile normalization was performed using "normalize.quantiles" function in "preprocessCore" R library (Bolstad, 2001).

2.3 PCA Analysis

Aside from technical variation, there can be biological variations that can interfere with results. For example, one individual might have had a different disease background, which could lead to dissimilar gene expression profile for that individual from the rest of the samples. Principle Component Analysis (PCA) is an efficient method to identify such differences. PCA is a dimension reduction technique which uses orthogonal transformation to convert multiple variables into linearly uncorrelated principle components, which are ordered by the amount of variance they explain (e.g. the first principle component explains the largest possible variance). PCA is a common method to analyse gene expression data since it is easier to represent each sample by a few principle components instead of using thousands of variables (number of genes).

I used the built-in R function "prcomp" with the "scale" argument to calculate prin-

ciple components (PCs). Then, I plotted first and second PCs against each other to visually inspect and identify outlier samples (**Figure 2.1 and Figure 2.2**).

According to the PCA analyses the following samples were removed from analysis:

- Liang2007: *"GSM119676"*, *"GSM119666"*
- Blalock2004: *"GSM21205"*, *"GSM21207"*

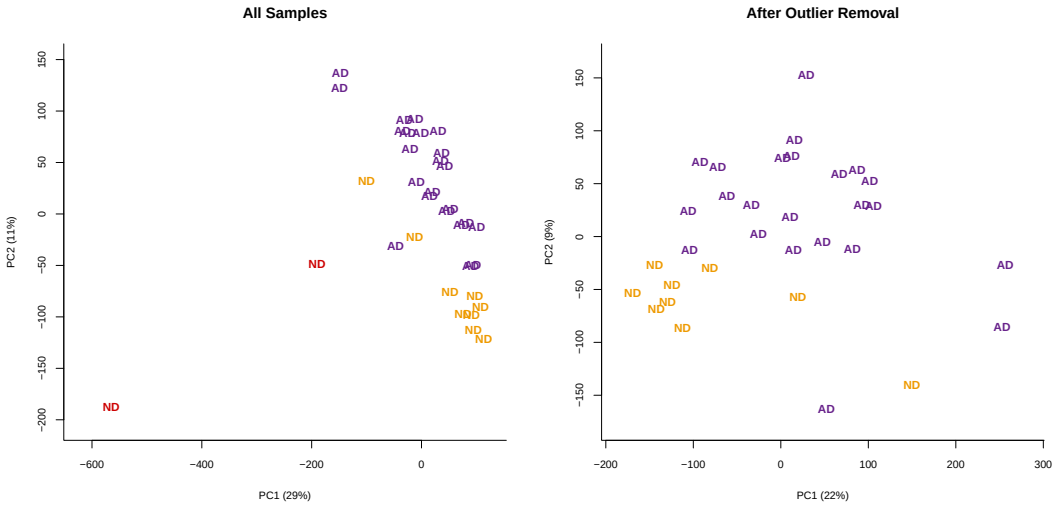

Figure 2.1: PCA analysis of the Liang2007 dataset before and after outlier removal. AD is for Alzheimer's Disease and ND is for not-dementia (control). Red coloured samples were identified as outliers and removed.

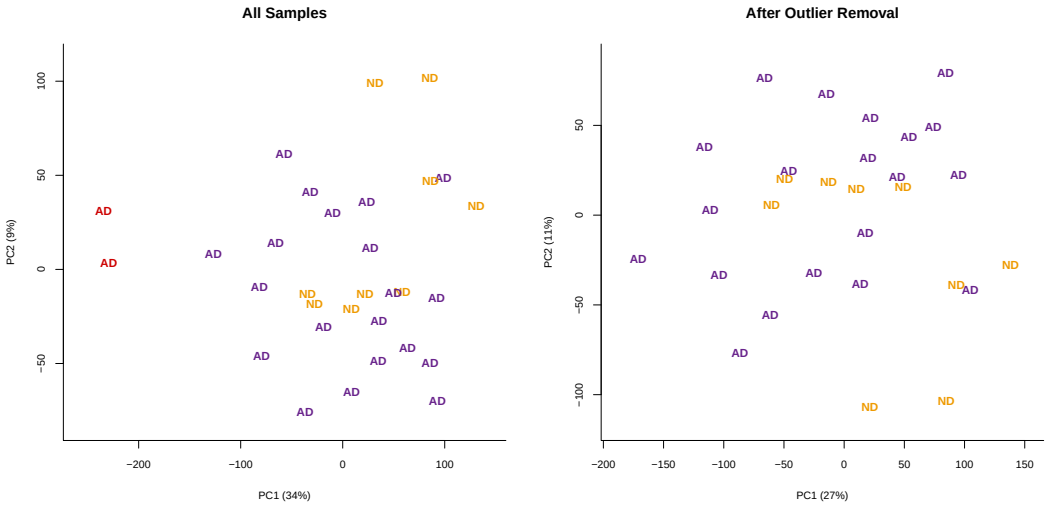

Figure 2.2: PCA analysis of Blalock2004 dataset before and after outlier removal. AD is for Alzheimer's Disease and ND is for not-dementia (control). Red coloured samples were identified as outliers and removed.

2.4 Differential Expression Test

To identify differentially expressed genes in Alzheimer's Disease (AD), I used the non-parametric Spearman's rank correlation test. The advantage of using a non-parametric test is that it is not affected by outliers, and does not assume a bivariate normal distribution. The correlation test statistic also facilitates comparison of AD-related changes with age-related expression changes (which is not part of this study, but will be conducted in the future). I used Spearman's rho value as a measure of differential expression under AD, i.e. to assess a gene whether its expression increased/decreased in AD compared to control. The range of rho value can be between -1 and 1 and its absolute magnitude shows the strength of association. If the rho value is positive, it means the gene expression increased in AD. Contrarily, if rho value is negative, it means the gene expression decreased in AD.

Each dataset has comparable number of AD and age-matched control samples for proper comparison. I used these two groups to calculate Spearman's rank correlation rho for each gene using the "cor.test" built-in R function, with the "method = 'spearman'" argument.

Genes differentially expressed in AD were identified using two different approaches. (1) Using the Spearman rank correlation test p-values, corrected by the Benjamini & Yekutieli method for multiple testing (**See 2.4.1**), and (2) identifying genes changing in the same direction, either increasing or decreasing in all datasets, and using a permutation test to assess significance and the false discovery rate (**See 2.6**).

2.4.1 Multiple Test Correction

Dealing with the results of thousands of statistical tests in each dataset brings with it the accumulation of false positives. This is known as the multiple comparison or multiple testing problem. The false discovery rate is the inferred proportion of false positives among all significant results. To overcome the effect of accumulating false positives due to multiple testing, p-values should be adjusted to keep false discovery rate (FDR) under control. There are several methods to control the false discovery rate. Benjamini and Hochberg (BH) method is the standard method to control the FDR. However, BH procedure loses FDR control with increasing positive dependence. The more conservative one is Benjamini and Yekutieli (BY) method, which is a two stage procedure, and can control FDR even with increasing positive dependence (Benjamini & Yekutieli, 2001). Therefore, I used the “Benjamini & Yekutieli" method to adjust p-values with “p.adjust” built-in R function using the “BY” argument. I applied a q<0.1 cutoff to the resulting values.

I performed multiple testing correction to Spearman’s rank correlation p-values to find differentially expressed genes in each dataset. As I explain in following Results section, there were few, if any genes passing multiple testing correction in most datasets. I then continued with the second approach to find shared differentially expressed genes across all datasets (**See 2.6**).

2.5 Correlation Across Datasets and Data Selection

To assess the correlation across datasets and brain regions, I used the Spearman’s rank correlation test without using any significance cutoff. Each dataset was compared to all others in pairwise manner. By this way, common genes only between

the paired datasets were considered. For these comparisons, I used the Spearman's rank correlation using "cor" R built-in function with the argument "spearman".

I used the "corrplot.mixed" function in the R "corrplot" library to visualize pairwise correlations across datasets. One dataset, Durrenberger2011, was removed from further analysis since its correlations with all other datasets were very low compared to correlations among all other pairs (**See Section 3.2**).

Using another approach to compare gene expression correlations across datasets, PCA analysis was applied. I used Spearman rho values to calculate principle components using "prcomp" function with scaling argument.

2.6 Permutation Test

After performing multiple testing correction, there were datasets in which any of the genes could not pass the significance cutoff ($q < 0.1$). Therefore, I was unable to obtain common genes across all datasets that change in same direction using a significance cutoff. For this reason, I implemented an alternative approach, by determining genes that change in all AD datasets in the same direction. Further, I used a permutation test to determine the statistical significance of the result.

For those datasets having only one brain region and thus one sample from one individual, AD and control samples were mixed with the built-in R function "sample", such that each individual was assigned randomly to the AD or control groups, keeping the group sample size fixed. Then, the same differential expression test (Spearman correlation) was applied and the results recorded. This procedure was repeated 1000 times.

However, those datasets with multiple brain regions include the same individual's samples, from more than one brain region. It is important to keep this dependence of samples (a type of "individual effect"), when conducting the permutation test. Otherwise, the similarity we find among datasets (our alternative hypothesis) could be due to similarity of individuals, insteaf of AD-control differences. Therefore, for those datasets, the permutation test was applied the same way as explained above,

while making sure that if one individual has been assigned to the AD group in one brain region, it was also assigned to the AD group in other brain regions for that permutation.

This random permutation process was applied 1000 times to all datasets. In each permutation step, the number of genes changing in the same direction across all datasets (hereafter "consistent genes") were calculated. I considered increasing and decreasing genes separately and obtained two null distributions. The false discovery rate for the consistent gene number was calculated as:

$FDR = N_{observed} / N_{expected}$,

where $N_{expected}$ representing the random expectation for the number of consistent genes, calculated as the median of the null distribution based on permutations, and $N_{observed}$ is the observed number of consistent genes. The p-value was calculated by the number of permutations having the same or higher number of observations of real data divided by total number of permutations.

2.7 Functional Analysis

Functional analysis was performed using the Gene Ontology (GO) database (Ashburner et al., 2000). GO is a hierarchical database describing gene functions in three main categories; biological process, molecular function and cellular component. For the analysis, I used GO Biological Process (BP) categories downloaded from the GO database at 03/26/2016 (Dönertas, 2016). There were 631 and 580 consistent up- and down-genes, respectively. The number of these genes having functional annotation in GO BP categories were 600 and 545, respectively. I performed the Fisher's Exact Test (FET) for up- and down-genes by employing a cutoff to GO groups to ensure that each GO group contains minimum number of 10 genes. There were 1877 GO groups passing this cutoff. I used the "fisher.test" R function, using the following contingency table (**Table 2.3**). This test gives odds ratio (OR) and the associated p-value.

Here OR>1, indicates enrichment of GO groups in terms of up-genes (gene expression increasing in AD) while OR<1 indicates enrichment of GO groups in terms of down-genes (gene expression decreasing in AD). Statistical significance of these

values were calculated using FET. However, the multiple comparison problem arises here again since FET was applied thousands of times. Therefore, p-values obtained with FET for 1877 GO groups were adjusted using "p.adjust" function with "BY" method (**Section 2.4.1**). Using q<0.1 as significance cutoff, enriched GO groups for up- and down-genes were identified.

2.7.1 Summarization of GO Categories with REVIGO

The outcome of GO enrichment analysis can be a long list of GO categories, which are highly redundant and cumbersome to interpret. As a result of the enrichment test, I obtained 343 GO BP categories enriched in upregulated genes and 96 GO BP categories enriched in downregulated genes in AD. To overcome this problem, I used the REVIGO software that summarizes long lists of redundant GO categories based on their semantic similarity which means the degree of shared genes among categories (Supek et al., 2011). REVIGO reduces the redundancy within list of GO categories calculating "uniqueness" and "dispensability". Dispensability is a measure of semantic similarity of two categories while uniqueness represents the negative similarity of a category to all other categories. Similar categories are clustered together using these measures. To visualize the REVIGO result, I used "treemap" R package (Martijn, 2017). Each rectangle in treemap is a representative cluster of summarized GO categories. The size of the rectangles is defined by the uniqueness of the categories. Similar clusters having same colours are joined together to form superclusters.

Table 2.2: Number of Probesets for Each Platform and Final Number of Genes After Summarization Step. Initial Probeset is the initial number of probesets when expression matrix is created with "exprs" function. "with ENSG" is the number of probesets having corresponding Ensembl Gene IDs. ">1 ENSG" is the number of probesets corresponding to more than one Ensembl Gene ID. Please note that when ">1 ENSG" is subtracted from "with ENSG", the outcome does not match "Total Probeset" since multiple probesets can correspond to multiple Ensembl Gene IDs which may also overlap. "Total Probeset" is the number of probesets after removing probesets corresponding to more than one Ensembl Gene ID. "Total ENSG" is the number of Ensembl Gene IDs after taking mean of multiple probesets corresponding the same Ensembl Gene ID. For the second part of the table, I used the same approach. First, I converted probesets to Entrez Gene IDs and then Entrez IDs to Ensembl Gene IDs.

Dataset	**# PS**	**# PS with ENSG**	**#PS >1 ENSG**	**Final # PS**	**Total ENSG**
Liang2007	54675	40333	1671	38662	18163
Blalock2004	22283	19571	903	18668	11653
Hisayama2014	33297	27040	10296	23651	22962
Tan2012	22011	17392	5120	14540	14356
Antonell2015	33297	27040	10296	23651	22962
	# PS	**Entrez**	**# >1 ENSG**	–	**Total ENSG**
Miller2013	48803	27388	21	–	17782
Durrenberger2011	20589	18052	5	–	16305
Narayanan2014	38759	21299	18	–	17119
Zhang2013_PFC	39005	25787	26	–	17867
Zhang2013_VCX	37185	24923	26	–	17388
Zhang2013_CRB	39084	25839	26	–	17897

Table 2.3: Contingency Table for Fisher's Exact Test. GO-X is the one GO category to be tested. Other GOs includes the all other GO categories to be tested.

	Up-genes	**Down-genes**
GO-x	a	b
Other GOs	c(600-a)	d(545-b)

CHAPTER 3

RESULTS

3.1 Gene Expression Change in AD in Each Dataset

I downloaded and preprocessed 13 transcriptome datasets from 9 different publications, representing 8 different brain regions, as described in Methods (**Table 2.2**). Each contained expression data calculated from postmortem brain samples of individuals diagnosed with AD and from roughly age-matched controls, with total sample sizes ranging from 14 to 466. The number of genes quantified in each dataset ranged from ∼11,000 to ∼23,000 (**Figure 3.1**). The least number of genes, 11653, was in Blalock2004 dataset and the most number of genes was in Hisayama2014 dataset.

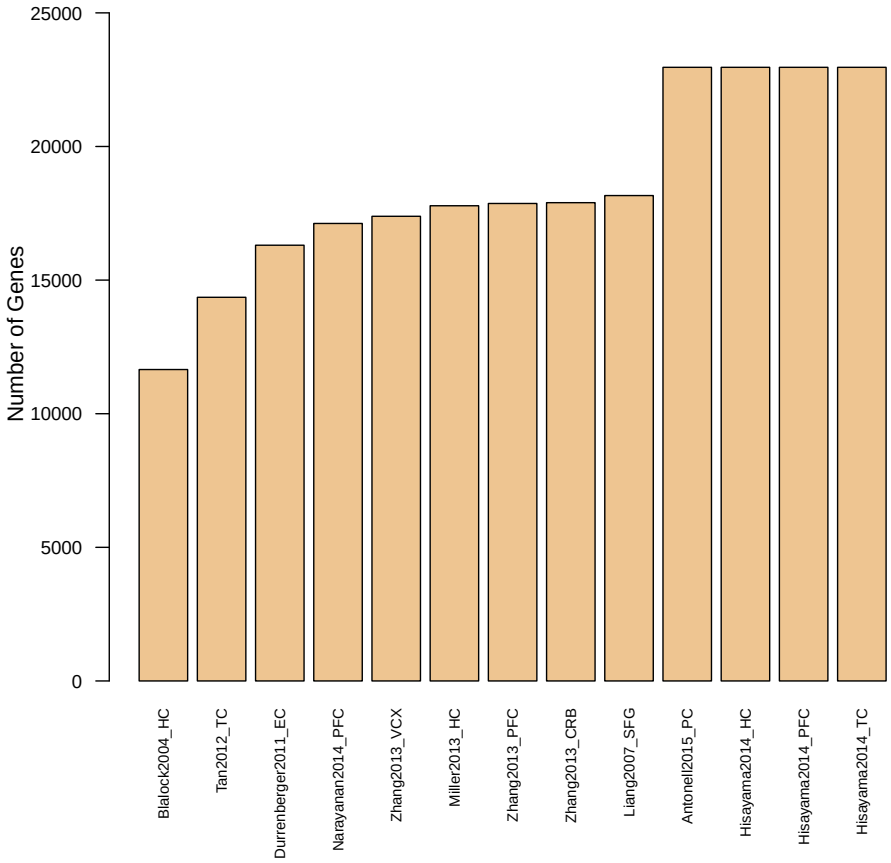

Figure 3.1: Number of Genes in Each Dataset

As I explain in Methods, I used alternative approaches to identify differentially expressed genes. First, I calculated the Spearman's rank correlation coefficient, or rho values of expression level change between two groups, AD and control, for each gene, in each dataset. The rho statistic ranges from -1 to 1, and here it is used as a measure of effect size, giving information about the strength and direction of gene expression change related to AD. I choose the first group being control and the second one AD, such that if the rho value is positive, that means the expression of a gene is increased in AD with respect to control, and vice versa. The higher the absolute value of rho, the more the expression of a gene has changed in AD compared to control. The significance of this change is given by p-value of this test. After doing multiple testing correction using the "Benjamini & Yekutieli" (BY) method, I calculated the proportion of genes showing significant change in AD,

using q<0.1 as cutoff. However, it was not surprising to find some datasets having no differentially expressed genes (no genes passing the cutoff), while others having thousands of differentially expressed genes (**Figure 3.2**). One reason is related to the sample size of the groups compared, AD and control; in other words, differences in statistical power among datasets. Datasets with small sample sizes had low or no differentially expressed genes, whereas datasets with large sample size (hundreds of samples in both groups) had high number of differentially expressed genes.

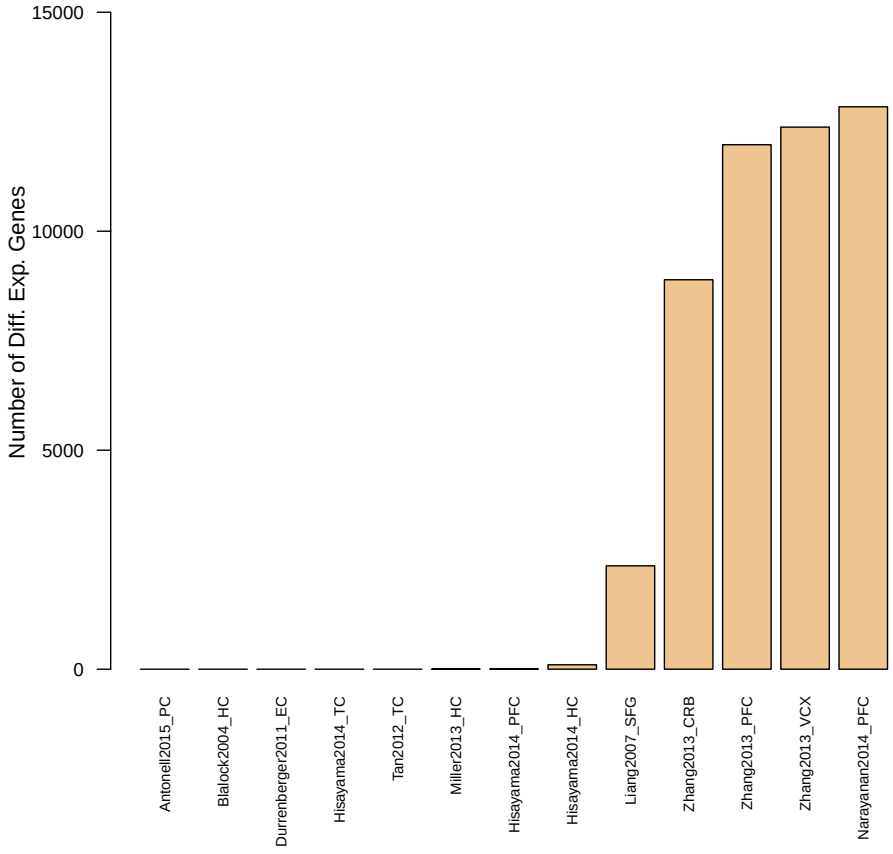

Figure 3.2: Number of differentially expressed genes in each dataset after multiple testing correction. Significance cutoff was set as q<0.1.

Second, I considered only the sign of rho values as an indicator of expression change, without employing a significance cutoff. I then calculated number of genes

showing upregulation and downregulation trends in AD. Out of 12 datasets, 10 datasets had more genes having increased expression in AD than genes with decreased expression (Wilcoxon signed rank test, p-value: 0.021) (**Figure 3.3**). This suggests that there is a tendency toward transcriptional upregulation in AD. However, whether the same genes across datasets change their expression in the same direction is not clarified by this approach.

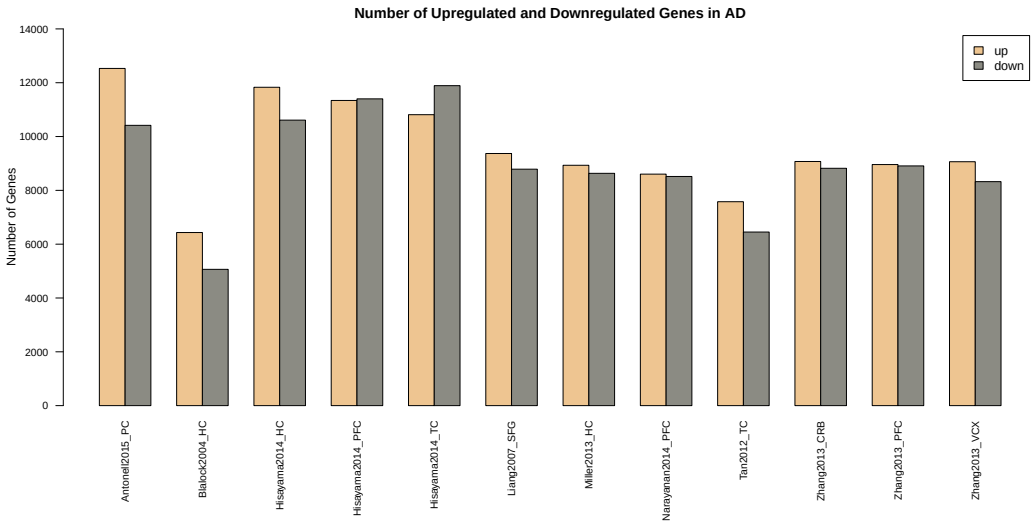

Figure 3.3: Gene expression change in AD compared to control. Legend: up: increased gene expression in AD compared to controls (i.e. rho value is positive), down: decreased gene expression in AD (i.e. rho value is negative).

3.2 Gene Expression Correlation Across Datasets

There were 13 datasets in total from 8 different brain regions. In order to assess the similarity among datasets in terms of AD-related gene expression change, I used the Spearman's rank correlation coefficient, or rho values calculated for each gene, as described above (**Section 3.1**). Then, between each dataset, across all overlapping genes, I calculated the pairwise correlation of correlation coefficients, and thus constructed the correlation matrix, again using Spearman's rank correlation.

Overall, the pairwise correlations between datasets were modest (**Figure 3.4**). The highest correlation was between prefrontal cortex (PFC) region of Zhang2013 dataset and PFC region of Narayanan2006 dataset (rho=0.98). Although it is tempting to claim that this might be due to the high similarity of transcription change between these two datasets, their similarity is more likely to be explained by the dominant effect of the platforms used, as Narayanan2006 and Zhang20013 datasets were both performed on Rosetta/Merck platform. Also, the correlation of visual cortex (VCX) region of Zhang2013 dataset with PFC region of Zhang2013 and PFC region of Narayanan2006 datasets was very high. This further supports the dominant platform effect for the similarity between these two datasets. Alternatively, Zhang2013 subdatasets (VCX and PFC) might also show high correlation due to their common laboratory of origin, and use of the same individuals. Interestingly, correlations among the three brain regions of the Hisayama2014 dataset are not as high as the ones among Zhang2013 subdatasets.

Correlation between AD-related expression change rho values in the Durrenberger2011 dataset with those from all other datasets was very low. The maximum correlation coefficient was $\sim$0.2 with the Narayanan2006 and Zhang2013 datasets, and went down to 0.1 (**Figure 3.4**). Overall, these numbers are conspiciously lower than the correlations among other datasets. There can be three possible reasons for this:

1. The Durrenberger2011 dataset represents AD-related change in the EC brain region, and it is the only dataset representing this region. It is possible that EC might have different transcriptome profile change in AD than other regions.

2. There may be technical issues related to this dataset. Experimental conditions might be different that might have led to technical bias, resulting in low correlation with other datasets in terms of gene expression change in AD.

3. Microarray platforms of datasets analysed in this study are not all the same. Durrenberger2011 and Miller2013 datasets were produced using the Illumina Beadchip platform. The gene expression distribution of these two datasets were more strongly right skewed than the other datasets (data not shown). It

is possible that the platform effect led to decrease in correlation with other datasets.

Because of one or a combinatorial effect of these possible reasons, I decided to exclude Durrenberger2011 from the analysis, which aims to identify convergent genes and pathways affected by AD.

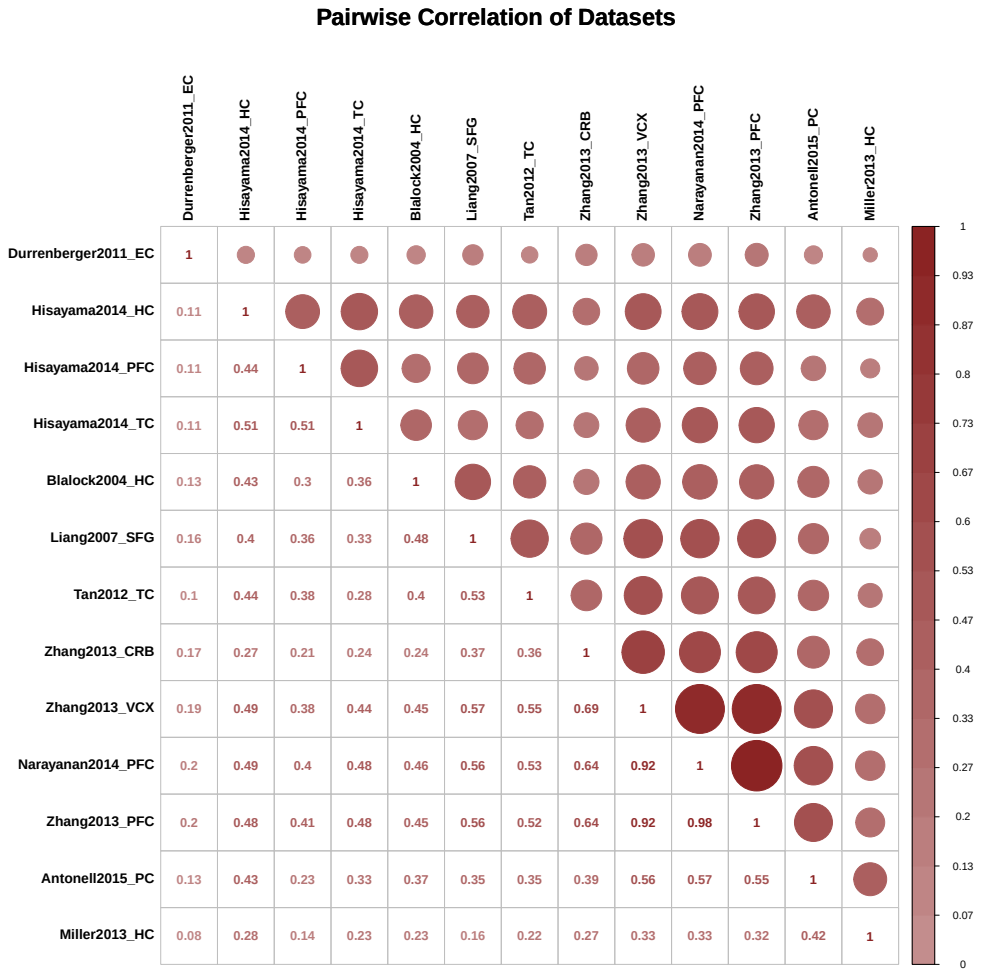

Figure 3.4: Correlation plot for gene expression change estimates (Spearman's rho values) in AD among all brain regions. Upper and lower panels are the different representation of the same result. The lower panel shows the exact Spearman correlation of correlations between two datasets, across all overlapping genes (min: 9535, max: 22962 overlapping genes). Upper panel shows the same result but using circles. The size of the circles and the density of the colour change with the magnitude of the correlation coefficient between two datasets. The datasets are ordered using hierarchical clustering of correlation coefficients between datasets.

To further investigate the brain region and platform effect on the AD-related expression change estimates among datasets, I performed PCA analysis with the 8704 common genes among all datasets, using the AD-related expression change (Spearman's rho) values per gene (not expression levels) of Spearman correlation test. In general, expression level distributions are strongly influenced by microarray platform type (data not shown). But, as the PCA plot in **Figure 3.5** shows, the type of microarray platform appaeared to have only modest effect on AD-related expression change estimates. More interestingly, we could observe no clustering among the datasets according to brain regions. According to the PCA, for example, a PFC dataset can show a more similar transriptome-wide AD response with a hippocampus (HC) dataset, than with another PFC dataset.

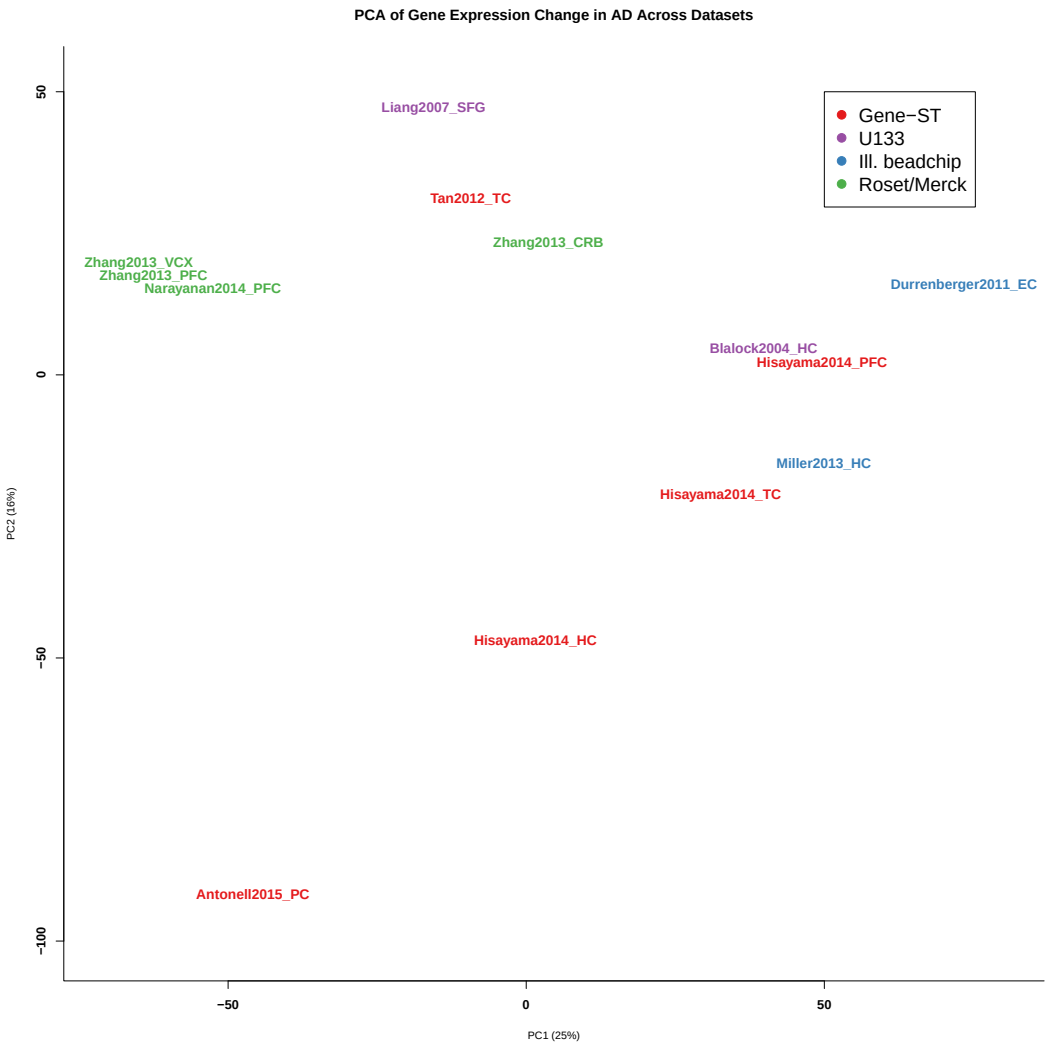

Figure 3.5: PCA of brain regions based on gene expression change in AD. Gene-ST is for HuGene-1_0-st, HuGene-1_1-st and HuEx-1_0-st platforms; U133 is for HG-U133_plus_2 and HG-U133A platforms; Ill. beadchip is for Illumina humanHT-12 V3.0 and Illumina humanRef-8 v2.0 beadchip platforms; Roset/Merck is for Human 44k 1.1 platform.

3.3 Common AD-Related Gene Expression Change Across Datasets

As the next step, I investigated common gene expression changes across datasets. For this purpose, I used Spearman's rank correlation rho value, as previously explained. As it was described in Section 3.1, after multiple testing correction with the BY method, no genes passing the significance cutoff q<0.1 were left, in half of datasets. Therefore, I did not employ any significance cutoff.

Instead, I concentrated on the set of common genes changing in the same direction in AD across all datasets. There were 631 such genes showing increased expression and 580 genes showing decreased expression in AD compared to controls. To test the significance of finding such number of common genes, I used a permutation scheme to establish the null distribution, representing the null hypothesis of no common AD effect among datasets. I thus randomized AD and control groups and calculated the Spearman's rho in each dataset, and recorded the common genes having increased/decreased expression in AD across all datasets in the permutations; I then used these numbers to construct the null distribution, representing 1000 random permutations. I then compared the observed values with the null distributions. As a result of this permutation test, the observed number of common genes were significant for both upregulated and downregulated genes among the 12 datasets (**Figure 3.5**). The full list of both gene groups are given in **Appendix A and B**.

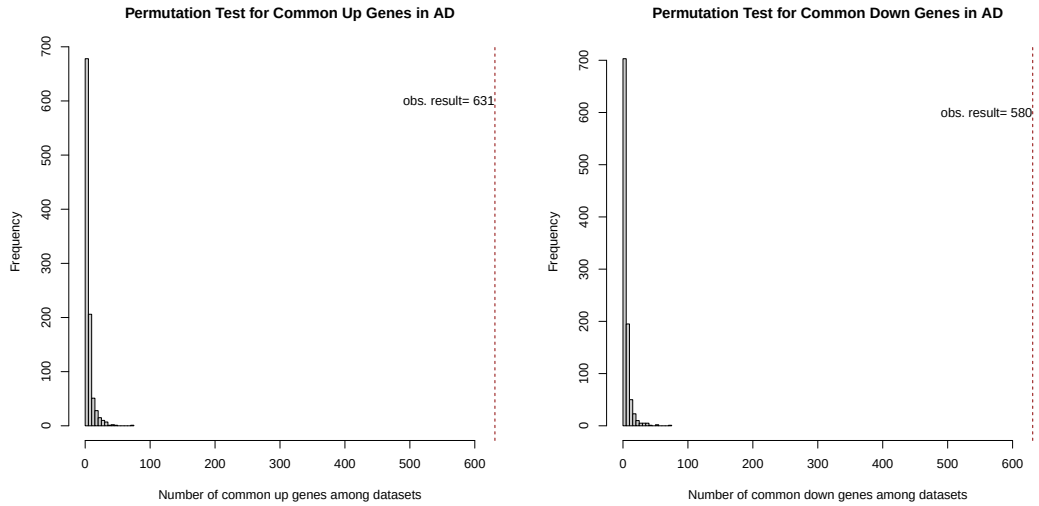

Figure 3.6: Permutation test for common upregulated/downregulated genes (up-genes and down-genes) across datasets. The left panel is for up-genes and right panel is for down-genes in AD. The figures show the distribution of the number of common genes among the 12 datasets in each permutation, which was done by randomizing AD and control groups in each dataset, calculating the Spearman's correlation rho for each gene, and determining if the rho values had the same sign across all 12 datasets. Dashed red lines show the observed results ("obs. result").

3.4 Functional Analysis of Gene Expression Change Related to AD

Using the Biological Process (BP) category of GO database, which is most easy to physiologically interpret, I performed functional analysis to find biological importance of AD-related gene expression change observed commonly among all brain regions. Previously, I found consistently 631 genes to be upregulated and 580 genes to be downregulated in AD (**Section 3.3**). I searched for the enrichment of upregulated genes compared to downregulated ones (odds ratio>1) and vice versa (odds ratio <1) in each GO Biological Process category. Then, I performed Fisher's exact test to assess the significance of enrichments for each category. To address the multiple testing problem, I adjusted the p-values using BY correction method and considered only significantly enriched GO categories (q<0.1).

3.4.1 GO BP Enrichment for Up Genes

As a result of the enrichment analysis, 343 GO Biological Process categories were found to be enriched in genes upregulated in AD compared to downregulated ones. The result was summarized using REVIGO interactive website algorithm (**Section 2.7.1**) and visualized as a treemap in (**Figure 3.7**). The full list of GO groups is given in the **Appendix C**.

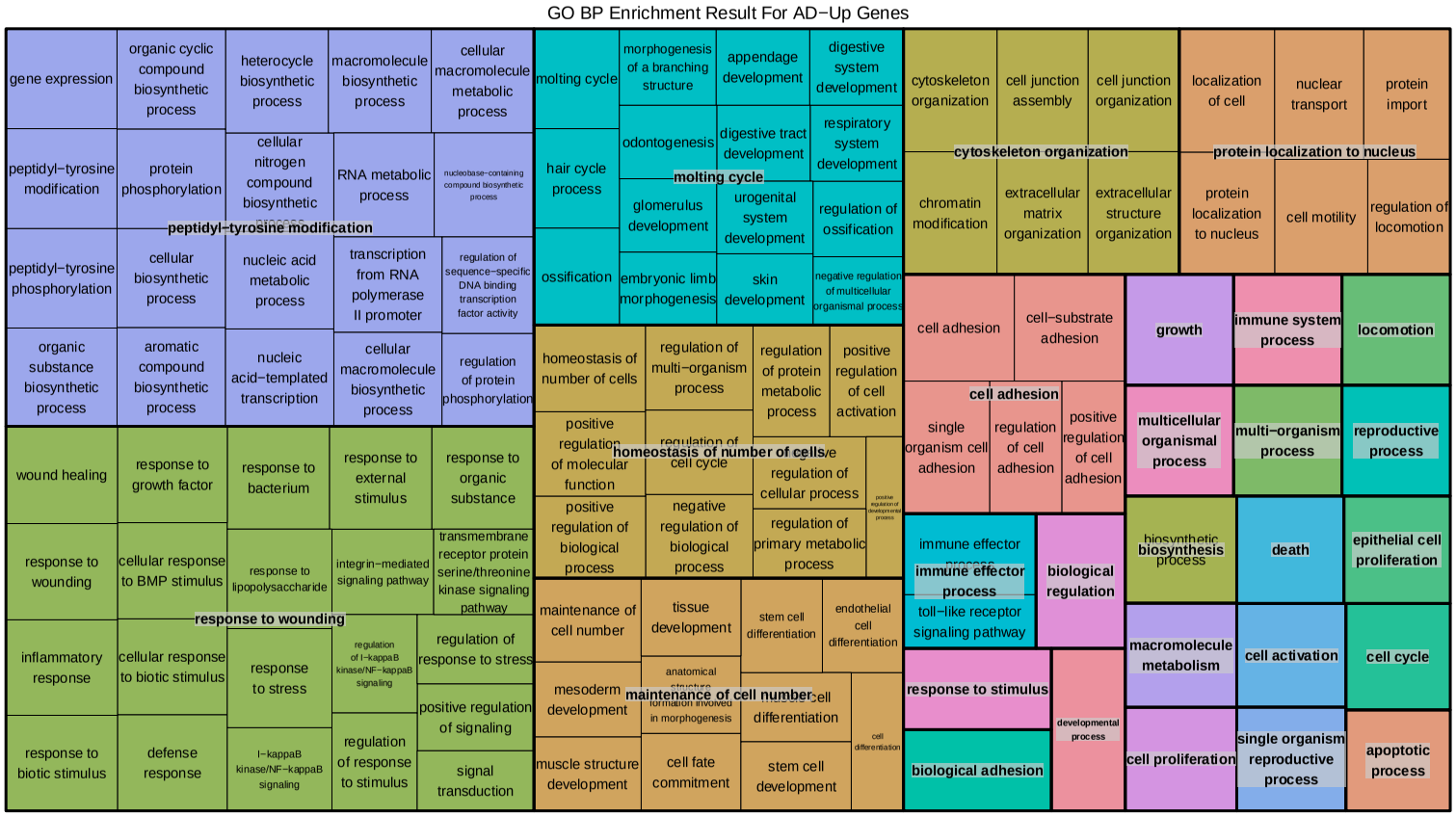

Figure 3.7: Biological Process enrichment result for the 600 common genes upregulated in AD across the 12 datasets. The significantly enriched GO groups (q<0.1) are summarized by REVIGO.

REVIGO summarized the 343 GO Biological Process enriched in upregulated common genes to 116 clusters, which were further joined into 7 superclusters of loosely related terms. It should be noted that the name of the supercluster is defined by the uniqueness of the GO categories in that supercluster. It does not necessarily represent the all GO categories. There weare mainly 7 superclusters and other small clusters summarized by REVIGO:

Peptydyl-tyrosine modification: This supercluster includes GO categories mostly involved in protein phosphorylation, nucleic acid metabolism, macromolecule biosynthetic process and RNA metabolic process. The most interesting GO categories in this supercluster are the ones associated with protein modification. The APP protein undergoes extensive post-translational modification and its deregulation results in amyloid-β deposition in AD. Moreover, the hyperphosphorylation of tau protein leads to the formation of NFT, which is the second hallmark of AD (**See Introduction**). Gene expression change in protein modification categories could cause disruptive alterations of regulation in post-translational mechanisms, which can be associated with amyloid-β and NFT pathogenesis.

Response to wounding: The GO categories in this supercluster mainly involve response to wounding/external stimulus/stress and signal transduction and its regulations. Abnormal accumulation of extracellular amyloid-β peptides and intracellular NFT disrupt cellular homeostatis and create cellular stress, which can result in expression change of stress-related genes. The "response to wounding" category contains several child terms including response to axon injury and wound healing. Although this category does not seem to be directly related to AD, it may be associated with damage response, as the neuronal cells might recognize NFTs as damage in cells. Neuronal cells also experience damage due to a variety of factors associated with oxidative stress, metabolic imbalances or genetic perturbations, which have accumulated throughout lifetime. These factors together might activate wound healing mechanisms. Several signal transduction pathways are also enriched in upregulated genes. For example, I-κB kinase signaling pathway is involved in inflammation. Together with other small clusters, e.g. "immune effector process" and "immune system process" (**Figure 3.7**), we see that immune system-related GO categories show enrichment in upregulated genes in AD. This result is consistent with previous findings that show upregulation of immune response genes

and regulatory regions (Bertram, Lill & Tanzi, 2010; Gjoneska et al., 2015)).

Molting cycle: All categories in this supercluster are developmental process-related categories including digestive system development, embryonic limb morphogenesis, skin development and ossification. These categories seem to be irrelevant to the brain tissue. However, it is possible that the regulators of these pathways overlap, and what we are observing is the outcome of an upregulation of these regulatory genes, resulting in the enrichment of developmental pathways. Several studies previously showed activation of mitotic proteins in AD (Arendt et al., 2000), including cell cycle markers, cyclin E (Nagy, Esiri, Cato & Smith, 1997), cyclin D, and cdk4 in the AD hippocampus region, but not in control subjects (Busser, Geldmacher & Herrup, 1998). It was hypothesized that neuronal cells attempt to re-enter the cell cycle, and induction of dedifferentiation leads to cell death possibly through apoptotic mechanisms. Indeed, supporting this hypothesis, I found the following GO categories enriched among upregulated genes in AD: cell cycle, cell proliferation, growth, death and apoptotic process.

Homeostasis of number of cells and maintenance of cell number: These two superclusters include noteworthy GO categories including stem cell differentiation, regulation of cell cycle, cell fate commitment, regulation of biological process, cell number maintenance and other development-related categories. Enrichment in differentiation and cell cycle-related categories further supports the above-explained hypothesis that neuronal cells might lose their terminally differentiated state and eventually die. Dividing cells have to adjust their number in tissue microenvironment to maintain homeostatis. Activation of cell cycle mechanisms in brain tissue can further initiate pathways related to these pathways, even though neurons actually will not divide. Also, the category "regulation of biological process" refers to many distinct cellular processes. Pathophysiological changes in AD and possible re-entry of neurons to cell cycle might trigger many biological processes.

Cytoskeleton organization: The tau protein is an important regulator of microtubules and hence affects cytoskeleton dynamics in neurons. Phosphorylation of tau protein is tightly regulated through dynamic activation/deactivation of tau kinases and tau phosphatases in normal neuronal cells (Gong & Iqbal, 2008). Aberrant upregulation of upstream genes, which are related to cytoskeleton organization,

due to various stochastic reasons, might lead to hyperphosphorylation of tau protein. Other GO categories in this supercluster are extracellular structure organization and cell junction organization. Enrichment in the extracellular organization category might be a cause or result of amyloid-β plaques. Expression change in extracellular matrix-related genes might contribute to deposition of amyloid-β protein. Alternatively, to overcompensate aberrant amyloid-β plaques in extracellular matrix, the neuronal cells might attempt to adapt to this impaired environment by changing expression of extracellular matrix-related genes. Lastly, enrichment in the category "cell junction organization and assembly" is explained below, together with cell adhesion-related GO categories.

Cell adhesion: The GO categories in this supercluster are all related to cell adhesion and its regulation. Several studies have established the relevance of cell adhesion molecules (CAMs) with AD pathology, especially with amyloid-β protein (Nielsen & Wennström, 2012). Synaptic CAMs interact with amyloid-β protein, as well as the enzymes involved in amyloid-β formation, which affects the expression and synaptic localization of CAMs (Leshchyns'Ka & Sytnyk, 2016). Upregulation of genes in these categories may change the cell junction organization and alter synaptic transmission, which will eventually result in degeneration of neuronal networks in AD.

Protein localization to nucleus: There are two main GO categories in this supercluster; one is about cell localization and the other one is about protein import into nucleus. The former one might seem to be irrelevant to a brain tissue. However, datasets I analyzed in this study are not from a single cell type, they represent a whole tissue. Also, microglia have been shown to have altered motility in AD (Gyoneva, Swanger, Zhang, Weinshenker & Traynelis, 2016). Enrichment of genes in GO categories related to locomotion and cell localization may represent immune cells activated in AD brain. The latter one includes protein localization to nucleus, nuclear transport and protein import. Intracellular alterations and pathological hallmarks of AD may lead to disruption of subcellular localization of proteins and alterations in nuclear transport proteins.

3.4.2 GO BP Enrichment for Down Genes

There are 94 GO Biological Process categories that are enriched in the 545 down-regulated genes in AD compared to upregulated ones. The result is summarized using the REVIGO algorithm and visualized in **Figure 3.8**. The full list of GO groups is given in **Appendix D**.

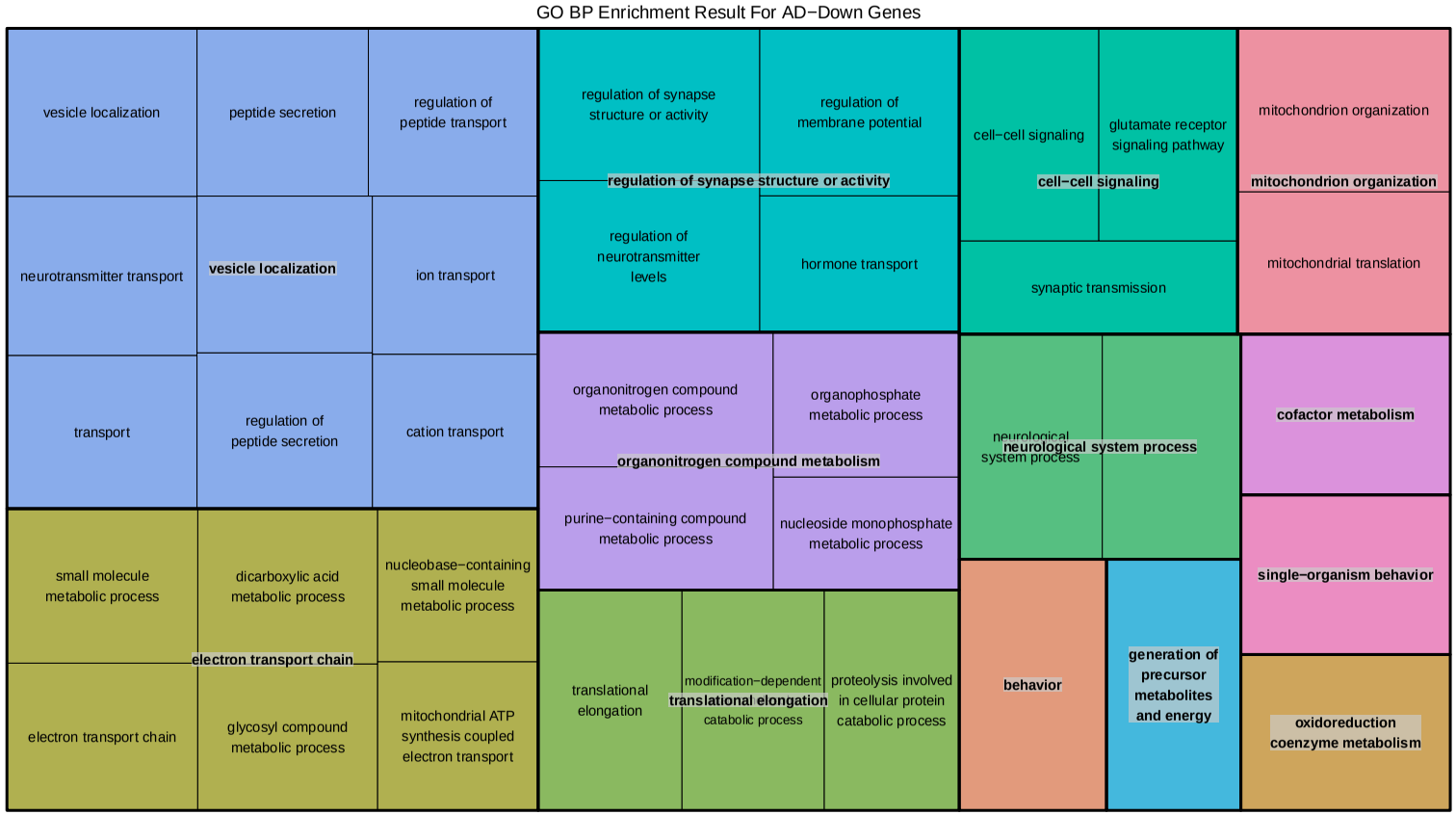

Figure 3.8: GO Biological Process enrichment result for the 545 common genes downregulated in AD. The 94 GO groups are summarized by REVIGO.

REVIGO summarized the 94 GO Biological Process categories to into 38 clusters, which were further joint into superclusters of loosely related terms. There are mainly 8 superclusters and other small clusters summarized by REVIGO:

Vesicle localization: The GO categories in this supercluster are related to cellular trafficking and vesicular transport. Vesicle localization is important in normal neuronal function. Ion transport and cation transport are related to action potential in neuronal cells. Neurotransmitter transport mechanism is an important feature of neurons to communicate with each other and establish neuronal networks. Peptides are also essential molecules secreted by neurons to communicate with each other. Downregulation of genes involved in these categories suggests impairment of neuronal function in AD.

Electron transport chain: This supercluster contains GO categories involved in energy metabolism. Previously, the downregulation of genes involved in mitochondrial energy metabolism has been shown **See Introduction**. Electron transport chain, mitochondrial ATP synthesis coupled electron transport categories are enriched among downregulated genes. These result supports the previous findings about alterations in energy metabolism in AD.

Mitochondrion organization: This supercluster contains two categories; mitochondrion organization and mitochondrial translation. Downregulation of genes in these categories also affects cellular energy metabolism.

Regulation of synapse structure or activity: The GO categories in this supercluster are related to synaptic transmission. Establishment of membrane potential is crucial for action potential. Neurotransmitters and synapse structure are self explanatory in their importance of synaptic transmission. Downregulation of genes in these categories may lead to synaptic loss and cognitive decline in AD.

Cell-cell signaling: This supercluster is also related to synaptic transmission and communication between cells. Downregulation of genes in these categories may contribute to impairments in communication between neurons and neuronal networks in AD.

Organonitrogen compound metabolism: Nitrogen and phosphate are important molecules in nucleotide metabolism. GO categories related to their metabolic process show downregulation in AD. Purine-containing compound metabolic process and nucleoside monophosphate metabolic process are enriched in downregulated genes. These results may indicate dysfunction in DNA and RNA biosynthetic processes in AD.

Translational elongation: This supercluster involves pathways related to translation and protein metabolic processes. Translational elongation contains several child terms including regulation of translation. Modification-dependent catabolic process and proteolysis involved in cellular protein catabolic process are other categories in this supercluster. Downregulation of genes in these categories may result in alterations in protein homeostasis from protein synthesis to proteolysis of proteins, which may contribute to AD pathogenesis.

Other categories enriched in downregulated genes are cofactor metabolism, single-organism behavior, oxidoreductase coenzyme metabolism, generation of precursor metabolites and energy and behavior.

CHAPTER 4

DISCUSSION

In this study, I investigated the gene expression changes associated with AD in different regions of the human brain. In total, I used 13 datasets, one of which was excluded from the analysis, leaving me with 12 datasets spanning 7 regions of the cerebral cortex, HC, SFG, TC, PFC, PC, and VCX, and also the cerebellum (CRB). I aimed to find common patterns associated with AD across these diverse brain regions, which are distinct both in their architecture, in their function, and in their known associations with AD progression.

First, I started analysis with raw datasets from Affymetrix microarrays. There were four different Affymetrix platforms, and three of them are commonly used platforms in microarray experiments: HuGene, HuEx, and HG-U133. I used publicly available R packages, "oligo" and "affy", to analyze datasets from these platforms. However, there was no freely available R package for the analysis of other two platforms; Illumina Beadchip and Rosetta/Merck. Therefore, I used the authors' preprocessed data, the "series matrix files" uploaded in the GEO database, to analyze these datasets (Miller2013, Durrenberger2011, Zhang2013 and Narayanan2006) from these platforms. One problem with using preprocessed data is that the analysis pipeline may not be the same across all datasets. I tried to compensate this drawback by normalising the preprocessed datasets in the same way as the others, using quantile normalization, at least to achieve the same normalisation in each datasets.

At the end of the preprocessing, I obtained gene annotations for each gene in each dataset. The Blalock2004 dataset had the lowest number of genes among all datasets. However, there were more than 10,000 genes in each of the datasets, which was enough to continue with the analysis. Number of genes across datasets ranged

from 11,000 to 23,000 (**Figure 3.1**). It is possible that the high variance in the number of genes among datasets can affect identification of common gene expression change patterns. In other words, the dataset with the fewest genes might be precluding identification of certain common expression change patterns. To address this problem, I performed the analysis excluding the dataset with the lowest gene number (Blalock2004). However, in the enrichment analysis, I found that the GO categories are extremely similar (data now shown).

Then, I studied gene expression change patterns in AD, in each dataset separately. For this I calculated the Spearman's rank correlation coefficient for expression change in AD vs. controls. Out of 12 datasets, 10 datasets had more genes upregulated in AD than downregulated ones (**Figure 3.3**). The difference appeared significant in a Wilcoxon signed rank test; however this approach does not take into account dependence among some of the datasets. I may in the future perform a permutation-based test to establish the significance of this observation. Taken at face value, this result may suggest that there is a trend for increased gene expression pattern in AD, and that this is shared across diverse brain regions.

As a next step, I calculated pairwise correlations between datasets using Spearman's rank correlation coefficient estimated for each gene in each dataset. This revealed a number of interesting observations: First, one dataset, Durrenberger2011, had particularly low correlations with all other datasets. The possible reasons for the very low correlation of this datasets were discussed in the Results section (**Result 3.2**): The cause could be biological, this being the only EC dataset, but it could also be technical. Assuming that a technical bias is more likely, in order to avoid unnecessary power loss in the analysis (since my aim is to find the common gene expression change patterns among datasets), I decided to exclude this dataset from the analysis.

Second, it was not possible to decide from the PCA analysis whether the type of microarray platform had a strong effect on AD-related gene expression change estimates. Thus, it is possible technical biases were limited for AD-related expression change calculations, and there is no prominent technical bias among datasets (perhaps Durrenberger2011 being an exception).

Third, and most importantly, the correlations between same brain regions were not higher than the other brain regions. For example, while the correlation of HC regions of Hisayama2014 and Blalock2004 datasets was 0.43, the correlation of HC regions of Hisayama2014 with PFC region of Narayanan2006 dataset was 0.49. In PCA analysis, it can likewise be seen that datasets do not group together according to the brain regions.

Next, I identified genes with common AD-related trends across all 12 datasets. Among over 8000 shared genes, there were 631 increasing and 580 decreasing genes in common. Since there was no significance cutoff, type II error rate is restrained by this way. To test the significance of these common genes, I used a random permutation procedure. Indeed, the result was highly significant for both upregulated ($p<0.001$) and downregulated ($p<0.001$) genes (**Figure 3.6**). Moreover, the advantage of using permutation approach to find common AD-related patterns across datasets is that the sample size variance among datasets is not a concerned issue.

As a next step, it is intriguing to ask whether these genes are enriched in particular GO BP categories. To test this, I performed enrichment analysis and found promising results. There were 343 GO BP categories enriched for upregulated genes and 94 GO BP categories enriched for downregulated genes. I summarized these results using REVIGO.

Among 343 GO categories enriched for upregulated genes, there are 7 main clusters, which appear generally highly relevant to pathophysiological changes in AD. The categories include protein modification pathways and extracellular matrix organisation. APP and tau proteins are the most important proteins in AD pathogenesis. Aberrant cleavage and post-translational modification of APP protein leads to accumulation of amyloid-β protein in extracellular matrix. Moreover, deregulated phosphorylation of tau protein results in intracellular neurofibrillary tangles (**See Introduction**). Enrichment of protein modification and extracellular matrix organisation categories in upregulated genes in AD may thus suggest a common dysfunction in pathways related to amyloid-β and tau protein modifications across brain regions.

Another interesting observation is that GO categories including cell cycle, apoptosis and death are enriched in upregulated genes in AD. Several studies have previously hypothesized the dedifferentiation of neurons and their possible attempt to re-entry to cell cycle in AD (Busser et al., 1998; Nagy et al., 1997) (**See also Section 3.4.1**). There are also GO categories enriched in stem cell differentiation, regulation of cell cycle and cell fate commitment in upregulated genes (**Figure 3.7**) while synaptic structure-related GO categories are enriched in downregulated genes (**See also Section 3.8**). It is tempting to argue that neuronal dedifferentiation might be a general phenomenon in AD, which leads neurons to possibly enter the cell-cycle, at least partly. Then, neurons can be arrested in cell cycle, which may lead to apoptosis. This phenomenon might explain the neuronal and synaptic loss in AD.

Notable GO categories enriched in downregulated genes in AD are related to neuron function, cell-cell signaling and cellular energy metabolism. Downregulation of genes in pathways related to ion transport, peptide secretion and neurotransmitter transport may impair communication between neurons. Moreover, a previous study has shown that energy metabolism related genes are downregulated predominantly in cingulate cortex region in AD brain (Liang2008). Possibly the same energy metabolism pathways may be downregulated in other regions as well, although perhaps not as dramatically as that in the cingulate cortex.

4.1 Limitations of the Study

1. Technical biases: I could not use the raw data for the analysis of datasets from the Rosetta/Merck and Illumina Beadchip arrays. Instead, I used preprocessed data from these arrays, which may introduce differences among datasets due to normalization and summarization methods. I tried to compensate this by normalizing these datasets using quantile normalization, but still, the expression level distributions from different platforms look distinct, strongly indicating the persistence of biases.

2. I excluded one dataset from this analysis, Durrenberger2011. It was the only dataset representing EC brain region. It was not possible to deduce whether very low correlation of this dataset with others is due to a technical bias

related to this dataset or that EC brain region has completely different transcription profile in AD. In the future, I could search for additional datasets representing this brain region to address this problem.

3. Some of the brain regions are represented only by one dataset in this analysis. I could extend the list of datasets to include in the analysis so that there should be multiple datasets representing each brain region.

4. All datasets I used in this analysis represent the late stages of AD. For some of the datasets, authors did not specify the stage of AD at all. The lack of exact AD stage specifications and the lack of early stage AD samples are another limitations of this study.

5. Except two datasets (Blalock2004 and Liang2007, which represent laser captured neuronal cell populations), all datasets are from whole tissue samples meaning that they represent transcription profiles not only from neurons but also other cells in the tissue including astrocytes and glia cells. I could perform cell type analysis to establish relative contribution of cell types ratio shifts to the transcription profile.

6. For the functional analysis, I only considered biological processes to investigate biological relevance of common genes across datasets. I could also include trans-regulators; miRNAs and transcription factors to find regulatory components of these genes.

7. All datasets included in this study are from microarray experiments. A more recent technology, RNA-seq, could be included in analysis. It could be possible to detect novel transcripts which might be attributed to AD. Also, the reproducibility between technical and biological replicates are higher in RNA-seq analysis which gives higher statistical power.

8. The common AD-related up and downregulated genes found in this study should be confirmed with other approaches such as experimental procedures. Genes showing the biggest changes according to their rho values could be confirmed using RT-PCT in a future study.

CHAPTER 5

CONCLUSION

Transcriptome studies on AD provide valuable information about the underlying molecular mechanisms of the disease. The availability of several published datasets focusing on AD-related gene expression changes in different brain regions have made it possible to perform meta-analyses, which can help both to increase sensitivity, to increase specificity, and to study common expression patterns among different brain regions. In this study, using published microarray datasets, I investigated the gene expression profiles from different brain regions affected by AD and searched for biological relevance of the common patterns among them. The outcomes of this study are as follows:

- There are hundreds of genes that are commonly affected by AD across all datasets and brain regions. I showed that we cannot expect to see these changes randomly by using permutation test. Interestingly, the effect of AD on hippocampus, frontal cortex and visual cortex converge into similar biological alterations suggesting that the differences in brain regions might be insignificant at least at this later stages of AD.

- Processes related to genes upregulated in AD include protein modification, protein localization, differentiation, cell cycle and apoptotic process. These findings point the dysfunction in protein modification pathways, which may affect the APP and tau protein homeostasis in neurons as well as the possible loss of differentiation and death of neurons.

- Processes related to genes downregulated in AD include cell-cell signaling, synaptic structure regulation and vesicle localisation. Communication between neurons is impaired by extracellular amyloid plaques while intracellular trafficking is affected by neurofibrillary tangles. These findings are

in line with the effects of pathological changes in AD and suggests that different brain regions share common pathways deregulated by AD.

REFERENCES

Alexander, G. E., Chen, K., Pietrini, P., Rapoport, S. I. & Reiman, E. M. (2002, may). Longitudinal PET evaluation of cerebral metabolic decline in dementia: A potential outcome measure in Alzheimer's disease treatment studies. *American Journal of Psychiatry*, *159*(5), 738–745. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11986126 doi: 10.1176/appi.ajp.159.5.738

Antonell, A., Lladó, A., Altirriba, J., Botta-Orfila, T., Balasa, M., Fernández, M., ... Shen, J. (2013, jul). A preliminary study of the whole-genome expression profile of sporadic and monogenic early-onset Alzheimer's disease. *Neurobiology of aging*, *34*(7), 1772–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23369545 doi: 10.1016/j.neurobiolaging.2012.12.026

Arendt, T., Holzer, M., Stöbe, A., Gärtner, U., Lüth, H. J., Brückner, M. K. & Ueberham, U. (2000). Activated mitogenic signaling induces a process of dedifferentiation in Alzheimer's disease that eventually results in cell death. *Annals of the New York Academy of Sciences*, *920*, 249–255. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11193159 doi: 10.1111/j.1749-6632.2000.tb06931.x

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... Sherlock, G. (2000, may). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10802651 doi: 10.1038/75556

Barnes, D. E. & Yaffe, K. (2011, sep). *The projected effect of risk factor reduction on Alzheimer's disease prevalence* (Vol. 10) (No. 9). NIH Public Access. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21775213 doi: 10.1016/S1474-4422(11)70072-2

Benjamini, Y. & Yekutieli, D. (2001). The Control of the False Discovery Rate in

Multiple Testing Under Dependency. *The Annals of Statistics*, *29*(4), 1165–1188.

Bertram, L., Lill, C. M. & Tanzi, R. E. (2010). The Genetics of Alzheimer Disease: Back to the Future. *Neuron*, *68*(2), 270–281. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19679070 doi: 10.1016/j.neuron.2010.10.013

Bird, T. D. (1993). *Alzheimer Disease Overview*. University of Washington, Seattle. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20301340

Blalock, E. M., Geddes, J. W., Chen, K. C., Porter, N. M., Markesbery, W. R. & Landfield, P. W. (2004, feb). Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(7), 2173–8. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=357071{&}tool=pmcentrez{&}rendertype=abstract doi: 10.1073/pnas.0308512100

Bolstad, B. (2001). Probe Level Quantile Normalization of High Density Oligonucleotide Array Data.

Braak, H. & Braak, E. (1991). Acta H ' pathologica Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol*, *82*(4), 239–259. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1759558 doi: 10.1007/BF00308809

Brewer, A. A. & Barton, B. (2014). Visual cortex in aging and Alzheimer's disease: changes in visual field maps and population receptive fields. *Frontiers in Psychology*, *5*(February), 74. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24570669 doi: 10.3389/fpsyg.2014.00074

Brickell, K. L., Steinbart, E. J., Rumbaugh, M., Payami, H., Schellenberg, G. D., Van Deerlin, V., ... Bird, T. D. (2006, sep). Early-onset Alzheimer disease in families with late-onset Alzheimer disease: a potential important subtype of familial Alzheimer disease. *Arch Neurol*, *63*(9), 1307–1311. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16966510 doi: 10.1001/archneur.63.9.1307

Busser, J., Geldmacher, D. S. & Herrup, K. (1998, apr). Ectopic cell cycle proteins predict the sites of neuronal cell death in Alzheimer's disease brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *18*(8), 2801–7. Retrieved from http://www.ncbi.nlm.nih.gov/

pubmed/9525997

Campion, D., Dumanchin, C., Hannequin, D., Dubois, B., Belliard, S., Puel, M., ... Frebourg, T. (1999, sep). Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum. *American journal of human genetics*, *65*(3), 664–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10441572 doi: 10.1086/302553

Carlson M. (n.d.). *Affymetrix Human Genome U133 Plus 2.0 Array annotation data (chip hgu133plus2). R package version 3.2.3.*

Carvalho, B. S. & Irizarry, R. A. (2010, oct). A framework for oligonucleotide microarray preprocessing. *Bioinformatics*, *26*(19), 2363–2367. Retrieved from https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btq431 doi: 10.1093/bioinformatics/btq431

Colangelo, V., Schurr, J., Ball, M. J., Pelaez, R. P., Bazan, N. G. & Lukiw, W. J. (2002, nov). Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: Transcription and neurotrophic factor down-regulation and up-regulation of apoptotic and pro-inflammatory signaling. *Journal of Neuroscience Research*, *70*(3), 462–473. Retrieved from http://doi.wiley.com/10.1002/jnr.10351 doi: 10.1002/jnr.10351

Corder, E. H., Saunders, A. M., Risch, N. J., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., ... Schmader, K. E. (1994, jun). Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nature genetics*, *7*(2), 180–184. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7920638 doi: 10.1038/ng0694-180

Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., ... Pericak-Vance, M. A. (1993, aug). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science (New York, N.Y.)*, *261*(5123), 921–923. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8346443 doi: 10.1126/science.8346443

Cruts, M. & Van Broeckhoven, C. (1998). Presenilin mutations in Alzheimer's disease. *Human Mutation*, *11*(3), 183–190. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9521418 doi: 10.1002/(SICI)1098-1004(1998)11:3<183::AID-HUMU1>3.0.CO;2-J

DeKosky, S. T. & Scheff, S. W. (1990, may). Synapse loss in frontal cortex biopsies in Alzheimer's disease: Correlation with cognitive severity. *Annals*

of Neurology, *27*(5), 457–464. Retrieved from http://doi.wiley.com/10.1002/ana.410270502 doi: 10.1002/ana.410270502

Dönertas, M. (2016). Meta-Analysis of Gene Expression Reversals in Ageing Brain.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A. & Huber, W. (2005, aug). BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics*, *21*(16), 3439–3440. Retrieved from https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bti525 doi: 10.1093/bioinformatics/bti525

Edgar, R., Domrachev, M. & Lash, A. E. (2002, jan). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*, *30*(1), 207–210. Retrieved from https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/30.1.207 doi: 10.1093/nar/30.1.207

Esch, F. S., Keim, P. S., Beattie, E. C., Blacher, R. W., Culwell, A. R., Oltersdorf, T., ... Ward, P. J. (1990, jun). Cleavage of amyloid beta peptide during constitutive processing of its precursor. *Science (New York, N.Y.)*, *248*(4959), 1122–1124. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2111583 doi: 10.1126/science.2111583

Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. (2004, feb). affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, *20*(3), 307–315. Retrieved from https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btg405 doi: 10.1093/bioinformatics/btg405

Giannakopoulos, P., Herrmann, F. R., Bussière, T., Bouras, C., Kövari, E., Perl, D. P., ... Hof, P. R. (2003, may). Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease. *Neurology*, *60*(9), 1495–500. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12743238 doi: 10.1212/01.WNL.0000063311.58879.01

Ginsberg, S. D., Hemby, S. E., Lee, V. M., Eberwine, J. H. & Trojanowski, J. Q. (2000, jul). Expression profile of transcripts in Alzheimer's disease tangle-bearing CA1 neurons. *Annals of neurology*, *48*(1), 77–87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10894219

Gjoneska, E., Pfenning, A. R., Mathys, H., Quon, G., Kundaje, A., Tsai, L.-H. & Kellis, M. (2015, feb). Conserved epigenomic signals in mice and hu-

mans reveal immune basis of Alzheimer's disease. *Nature*, *518*(7539), 365–369. Retrieved from http://dx.doi.org/10.1038/nature14252 doi: 10.1038/nature14252

Gong, C.-X. & Iqbal, K. (2008). Hyperphosphorylation of microtubule-associated protein tau: a promising therapeutic target for Alzheimer disease. *Current medicinal chemistry*, *15*(23), 2321–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18855662

Grehan, S., Tse, E. & Taylor, J. M. (2001, feb). Two distal downstream enhancers direct expression of the human apolipoprotein E gene to astrocytes in the brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *21*(3), 812–822. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11157067 doi: 21/3/812[pii]

Gyoneva, S., Swanger, S. A., Zhang, J., Weinshenker, D. & Traynelis, S. F. (2016, aug). Altered motility of plaque-associated microglia in a model of Alzheimer's disease. *Neuroscience*, *330*, 410–420. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27288150 doi: 10.1016/j.neuroscience.2016.05.061

Hampel, H., Bürger, K., Teipel, S. J., Bokde, A. L. W., Zetterberg, H. & Blennow, K. (2008, jan). *Core candidate neurochemical and imaging biomarkers of Alzheimer's disease* (Vol. 4) (No. 1). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18631949 doi: 10.1016/j.jalz.2007.08.006

Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M. L., ... Williams, J. (2009, oct). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature genetics*, *41*(10), 1088–93. Retrieved from http://www.nature.com/doifinder/10.1038/ng.440 doi: 10.1038/ng.440

Hérard, A. S., Besret, L., Dubois, A., Dauguet, J., Delzescaux, T., Hantraye, P., ... Moya, K. L. (2006, dec). siRNA targeted against amyloid precursor protein impairs synaptic activity in vivo. *Neurobiology of Aging*, *27*(12), 1740–1750. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16337035 doi: 10.1016/j.neurobiolaging.2005.10.020

Hippius, H. & Neundorfer, G. (2003, mar). The discovery of Alzheimer's disease. *Dialogues in Clinical Neuroscience*, *5*(1), 101–108. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22034141 doi: 10.1055/s-2008-1026558

Ho, A. & Südhof, T. C. (2004, feb). Binding of F-Spondin to Amyloid-β

Precursor Protein: A Candidate Amyloid-β Precursor Protein Ligand that Modulates Amyloid-β Precursor Protein Cleavage. *Proceedings of the National Academy of Sciences*, *101*(8), 2548–2553. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14983046 doi: 10.1073/pnas.0308655100

Hokama, M., Oka, S., Leon, J., Ninomiya, T., Honda, H., Sasaki, K., ... Nakabeppu, Y. (2014, sep). Altered expression of diabetes-related genes in Alzheimer's disease brains: The Hisayama study. *Cerebral Cortex*, *24*(9), 2476–2488. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23595620 doi: 10.1093/cercor/bht101

Huang, Y. (2006, sep). Molecular and cellular mechanisms of apolipoprotein E4 neurotoxicity and potential therapeutic strategies. *Current opinion in drug discovery & development*, *9*(5), 627–41. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17002223

Hung, A. Y., Koo, E. H., Haass, C. & Selkoe, D. J. (1992, oct). Increased expression of beta-amyloid precursor protein during neuronal differentiation is not accompanied by secretory cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(20), 9439–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1409654 doi: 10.1073/pnas.89.20.9439

Jarvik, G., Larson, E. B., Goddard, K., Schellenberg, G. D. & Wijsman, E. M. (1996, jan). Influence of apolipoprotein E genotype on the transmission of Alzheimer disease in a community-based sample. *American journal of human genetics*, *58*(1), 191–200. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8554056

Katzman, R., Katzman R, K. T., RD, N., Neumann MA, C. R., Tomlinson BE, Blessed G, R. M., Blessed G, Tomlinson BE, R. M., ... ER, P. (1976, apr). The Prevalence and Malignancy of Alzheimer Disease. *Archives of Neurology*, *33*(4), 217. Retrieved from http://archneur.jamanetwork.com/article.aspx?doi=10.1001/archneur.1976.00500040001001 doi: 10.1001/archneur.1976.00500040001001

Khachaturian, A. S., Corcoran, C. D., Mayer, L. S., Zandi, P. P. & Breitner, J. C. S. (2004, may). Apolipoprotein E epsilon4 count affects age at onset of Alzheimer disease, but not lifetime susceptibility: The Cache County Study. *Archives of general psychiatry*, *61*(5), 518–524. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15123497 doi: 10.1001/archpsyc.61.5.518

Lambert, J.-C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., ... Amouyel, P. (2009, oct). Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nature genetics*, *41*(10), 1094–1099. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19734903 doi: 10.1038/ng.439

Lammich, S., Kojro, E., Postina, R., Gilbert, S., Pfeiffer, R., Jasionowski, M., ... Fahrenholz, F. (1999, mar). Constitutive and regulated alpha-secretase cleavage of Alzheimer's amyloid precursor protein by a disintegrin metalloprotease. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(7), 3922–3927. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10097139 doi: 10.1073/pnas.96.7.3922

Lee, M. S., Kao, S. C., Lemere, C. A., Xia, W., Tseng, H. C., Zhou, Y., ... Tsai, L. H. (2003, oct). APP processing is regulated by cytoplasmic phosphorylation. *Journal of Cell Biology*, *163*(1), 83–95. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14557249 doi: 10.1083/jcb.200301115

Leshchyns'Ka, I. & Sytnyk, V. (2016). *Synaptic Cell Adhesion Molecules in Alzheimer's Disease* (Vol. 2016). Hindawi Publishing Corporation. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27242933 doi: 10.1155/2016/6427537

Leverenz, J. B. & Raskind, M. A. (1998, apr). Early amyloid deposition in the medial temporal lobe of young Down syndrome patients: a regional quantitative analysis. *Experimental neurology*, *150*(2), 296–304. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9527899 doi: 10.1006/exnr.1997.6777

Leyssen, M., Ayaz, D., Hébert, S. S., Reeve, S., De Strooper, B. & Hassan, B. A. (2005, aug). Amyloid precursor protein promotes post-developmental neurite arborization in the Drosophila brain. *The EMBO journal*, *24*(16), 2944–2955. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16052209 doi: 10.1038/sj.emboj.7600757

Liang, W. S., Reiman, E. M., Valla, J., Dunckley, T., Beach, T. G., Grover, A., ... Stephan, D. A. (2008, mar). Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(11), 4441–6. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2393743{&}tool=pmcentrez{&}rendertype=abstract doi: 10.1073/pnas

.0709259105

Loring, J. F., Wen, X., Lee, J. M., Seilhamer, J. & Somogyi, R. (2001). A gene expression profile of Alzheimer's disease. *DNA & Cell Biology*, *20*(11), 683–695.

Lott, I. T. & Head, E. (2001, aug). *Down syndrome and alzheimer's disease: A link between development and aging* (Vol. 7) (No. 3). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11553933http://doi.wiley.com/10.1002/mrdd.1025 doi: 10.1002/mrdd.1025

Ma, J., Yee, A., Brewer, H. B., Das, S. & Potter, H. (1994, nov). *Amyloid-associated proteins α1-antichymotrypsin and apolipoprotein E promote assembly of Alzheimer β-protein into filaments* (Vol. 372) (No. 6501). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7969426 doi: 10.1038/372092a0

Mahley, R. W., Weisgraber, K. H. & Huang, Y. (2006, apr). Apolipoprotein E4: a causative factor and therapeutic target in neuropathology, including Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(15), 5644–51. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16567625 doi: 10.1073/pnas.0600549103

Martijn, T. (2017). Treemap Visualization.

Matsui, T., Ingelsson, M., Fukumoto, H., Ramasamy, K., Kowa, H., Frosch, M. P., ... Hyman, B. T. (2007, aug). Expression of APP pathway mRNAs and proteins in Alzheimer's disease. *Brain Research*, *1161*(1), 116–123. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17586478 doi: 10.1016/j.brainres.2007.05.050

Mayeux, R., Saunders, A. M., Shea, S., Mirra, S., Evans, D., Roses, A. D., ... Phelps, C. H. (1998, feb). Utility of the apolipoprotein E genotype in the diagnosis of Alzheimer's disease. Alzheimer's Disease Centers Consortium on Apolipoprotein E and Alzheimer's Disease. *The New England journal of medicine*, *338*(8), 506–511. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9468467 doi: 10.1056/NEJM199802193380804

Miller, J. A., Woltjer, R. L., Goodenbour, J. M. & Horvath, S. (2013). Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome medicine*, *5*(5), 48. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23705665 doi: 10.1186/gm452

Nagy, Z., Esiri, M. M., Cato, A. M. & Smith, A. D. (1997, jul). Cell cycle markers

in the hippocampus in Alzheimer's disease. *Acta neuropathologica*, *94*(1), 6–15. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9224524

Narayanan, M., Huynh, J. L., Wang, K., Yang, X., Yoo, S., McElwee, J., ... Zhu, J. (2014, jan). Common dysregulation network in the human prefrontal cortex underlies two neurodegenerative diseases. *Molecular systems biology*, *10*, 743. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4299500{&}tool=pmcentrez{&}rendertype=abstract

Nielsen, H. & Wennström, H. M. (2012, jul). Cell adhesion molecules in Alzheimer's disease. *Degenerative Neurological and Neuromuscular Disease*, *Volume 2*, 65. Retrieved from http://www.dovepress.com/cell-adhesion-molecules-in-alzheimer39s-disease-peer-reviewed-article-DNND doi: 10.2147/DNND.S19829

O'Brien, R. J. & Wong, P. C. (2010). Amyloid Precursor Protein Processing and Alzheimers Disease. *Annual review of neuroscience*, *34*(March), 183–202. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21456963 doi: 10.1146/annurev-neuro-061010-113613

Perez, R. G., Zheng, H., Van der Ploeg, L. H. & Koo, E. H. (1997, dec). The beta-amyloid precursor protein of Alzheimer's disease enhances neuron viability and modulates neuronal polarity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *17*(24), 9407–9414. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9390996

Prasher, V. P., Farrer, M. J., Kessling, A. M., Fisher, E. M. C., West, R. J., Barber, P. C. & Butler, A. C. (1998, mar). Molecular mapping of Alzheimer-type dementia in Down's syndrome. *Annals of Neurology*, *43*(3), 380–383. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9506555http://doi.wiley.com/10.1002/ana.410430316 doi: 10.1002/ana.410430316

Puthiyedth, N., Riveros, C., Berretta, R. & Moscato, P. (2016). Identification of differentially expressed genes through integrated study of Alzheimer's disease affected brain regions. *PLoS ONE*, *11*(4), e0152342. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27050411http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4822961 doi: 10.1371/journal.pone.0152342

Saunders, A. M., Strittmatter, W. J., Schmechel, D., George-Hyslop, P. H., Pericak-Vance, M. A., Joo, S. H., ... Alberts, M. J. (1993, aug). Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic

Alzheimer's disease. *Neurology*, *43*(8), 1467–1472. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8350998 doi: 10.1212/WNL.43.8.1467

Selkoe, D. J. (1998, nov). The cell biology of beta-amyloid precursor protein and presenilin in Alzheimer's disease. *Trends in cell biology*, *8*(11), 447–53. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9854312 doi: 10.1016/S0962-8924(98)01363-4

Seshadri, S., Fitzpatrick, A. L., Ikram, M. A., DeStefano, A. L., Gudnason, V., Boada, M., ... EADI1 Consortium (2010, may). Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA*, *303*(18), 1832–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20460622 doi: 10.1001/jama.2010.574

Sherrington, R., Rogaev, E. I., Liang, Y., Rogaeva, E. A., Levesque, G., Ikeda, M., ... St George-Hyslop, P. H. (1995, jun). Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature*, *375*(6534), 754–60. Retrieved from http://www.nature.com/doifinder/10.1038/375754a0 doi: 10.1038/375754a0

Soba, P., Eggert, S., Wagner, K., Zentgraf, H., Siehl, K., Kreger, S., ... Beyreuther, K. (2005, oct). Homo- and heterodimerization of APP family members promotes intercellular adhesion. *The EMBO journal*, *24*(20), 3624–3634. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16193067 doi: 10.1038/sj.emboj.7600956

Sun, Y., Wu, S., Bu, G., Onifade, M. K., Patel, S. N., LaDu, M. J., ... Holtzman, D. M. (1998, may). Glial fibrillary acidic protein-apolipoprotein E (apoE) transgenic mice: astrocyte-specific expression and differing biological effects of astrocyte-secreted apoE3 and apoE4 lipoproteins. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *18*(9), 3261–3272. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9547235

Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., Rivals, I., Personnaz, L., ... Törönen, P. (2011, jul). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, *6*(7), e21800. Retrieved from http://dx.plos.org/10.1371/journal.pone.0021800 doi: 10.1371/journal.pone.0021800

Tan, M. G., Chua, W. T., Esiri, M. M., Smith, A. D., Vinters, H. V. & Lai, M. K. (2010). Genome wide profiling of altered gene expression in the neocortex of Alzheimer's disease. *Journal of Neuroscience Research*, *88*(6), 1157–

1169. Retrieved from http://doi.wiley.com/10.1002/jnr.22290 doi: 10.1002/jnr.22290

Tharp, W. G. & Sarkar, I. N. (2013, apr). Origins of amyloid-β. *BMC Genomics*, *14*(1), 290. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23627794 doi: 10.1186/1471-2164-14-290

Van Cauwenberghe, C., Van Broeckhoven, C. & Sleegers, K. (2016, may). The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genetics in medicine : official journal of the American College of Medical Genetics*, *18*(5), 421–430. Retrieved from http://www.nature.com/doifinder/10.1038/gim.2015.117 doi: 10.1038/gim.2015.117

Van den Heuvel, C., Blumbergs, P. C., Finnie, J. W., Manavis, J., Jones, N. R., Reilly, P. L. & Pereira, R. A. (1999, oct). Upregulation of amyloid precursor protein messenger RNA in response to traumatic brain injury: an ovine head impact model. *Experimental neurology*, *159*(2), 441–450. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10506515 doi: 10.1006/exnr.1999.7150

Vetrivel, K. S., Zhang, Y.-w., Xu, H. & Thinakaran, G. (2006, jun). Pathological and physiological functions of presenilins. *Molecular neurodegeneration*, *1*, 4. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16930451 doi: 10.1186/1750-1326-1-4

Wang, B., Yang, W., Wen, W., Sun, J., Su, B., Liu, B., ... Zhang, X. (2010, nov). γ-Secretase Gene Mutations in Familial Acne Inversa. *Science*, *330*(6007), 1065–1065. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20929727 doi: 10.1126/science.1196284

Weingarten, M. D., Lockwood, A. H., Hwo, S. Y. & Kirschner, M. W. (1975, may). A protein factor essential for microtubule assembly. *Proceedings of the National Academy of Sciences of the United States of America*, *72*(5), 1858–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1057175 doi: 10.1073/pnas.72.5.1858

Wruck, W., Schröter, F. & Adjaye, J. (2016, feb). Meta-Analysis of Transcriptome Data Related to Hippocampus Biopsies and iPSC-Derived Neuronal Cells from Alzheimer's Disease Patients Reveals an Association with FOXA1 and FOXA2 Gene Regulatory Networks. *Journal of Alzheimer's Disease*, *50*(4), 1065–1082. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26890743http://www.medra.org/

servlet/aliasResolver?alias=iospress{&}doi=10.3233/JAD-150733 doi: 10.3233/JAD-150733

Xu, Q., Bernardo, A., Walker, D., Kanegawa, T., Mahley, R. W. & Huang, Y. (2006, may). Profile and regulation of apolipoprotein E (ApoE) expression in the CNS in mice with targeting of green fluorescent protein gene to the ApoE locus. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *26*(19), 4985–94. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16687490 doi: 10.1523/JNEUROSCI.5476-05.2006

Yamazaki, T., Koo, E. H. & Selkoe, D. J. (1997, feb). Cell surface amyloid beta-protein precursor colocalizes with beta 1 integrins at substrate contact sites in neural cells. *The Journal of Neuroscience*, *17*(3), 1004–1010. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8994055

Yoshikai, S.-i., Sasaki, H., Doh-ura, K., Furuya, H. & Sakaki, Y. (1990, mar). Genomic organization of the human amyloid beta-protein precursor gene. *Gene*, *87*(2), 257–263. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2110105 doi: 10.1016/0378-1119(90)90310-N

Zhang, B., Gaiteri, C., Bodea, L. G., Wang, Z., McElwee, J., Podtelezhnikov, A. A., ... Emilsson, V. (2013, apr). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*, *153*(3), 707–720. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23622250 doi: 10.1016/j.cell.2013.03.030

APPENDIX A

LIST OF AD-RELATED GENES SHOWING UPREGULATION TREND

Table A.1: List of genes showing significant upregulation trend across datasets. *M-rho* is the median rho value of a gene among datasets.

ENSG ID	**Gene Name**	***M-rho***
ENSG00000143772	ITPKB	0.672
ENSG00000176046	NUPR1	0.663
ENSG00000173039	RELA	0.642
ENSG00000119950	MXI1	0.64
ENSG00000138193	PLCE1	0.637
ENSG00000183864	TOB2	0.635
ENSG00000112851	ERBIN	0.632
ENSG00000173530	TNFRSF10D	0.63
ENSG00000133789	SWAP70	0.624
ENSG00000162909	CAPN2	0.621
ENSG00000003989	SLC7A2	0.619
ENSG00000129116	PALLD	0.615
ENSG00000119900	OGFRL1	0.614
ENSG00000067182	TNFRSF1A	0.608
ENSG00000134324	LPIN1	0.604
ENSG00000026508	CD44	0.604
ENSG00000125733	TRIP10	0.602
ENSG00000111783	RFX4	0.599
ENSG00000162889	MAPKAPK2	0.597
ENSG00000141232	TOB1	0.597

Table A.1 (continued)

ENSG00000106624	AEBP1	0.593
ENSG00000129675	ARHGEF6	0.589
ENSG00000111907	TPD52L1	0.586
ENSG00000134531	EMP1	0.586
ENSG00000105855	ITGB8	0.583
ENSG00000113916	BCL6	0.582
ENSG00000172493	AFF1	0.582
ENSG00000152661	GJA1	0.582
ENSG00000091436	AC013461.1	0.581
ENSG00000152137	HSPB8	0.58
ENSG00000116729	WLS	0.579
ENSG00000101849	TBL1X	0.578
ENSG00000183255	PTTG1IP	0.577
ENSG00000159176	CSRP1	0.577
ENSG00000125398	SOX9	0.576
ENSG00000168309	FAM107A	0.573
ENSG00000173926	MARCH3	0.572
ENSG00000111961	SASH1	0.571
ENSG00000170525	PFKFB3	0.571
ENSG00000135063	FAM189A2	0.57
ENSG00000069702	TGFBR3	0.567
ENSG00000148175	STOM	0.565
ENSG00000175215	CTDSP2	0.564
ENSG00000130254	SAFB2	0.564
ENSG00000164949	GEM	0.563
ENSG00000137693	YAP1	0.562
ENSG00000131626	PPFIA1	0.561
ENSG00000137198	GMPR	0.559
ENSG00000060138	YBX3	0.559
ENSG00000144909	OSBPL11	0.554
ENSG00000174306	ZHX3	0.554
ENSG00000010810	FYN	0.552
ENSG00000164877	MICALL2	0.551
ENSG00000099875	MKNK2	0.551
ENSG00000065526	SPEN	0.549
ENSG00000124942	AHNAK	0.549
ENSG00000172380	GNG12	0.549
ENSG00000164050	PLXNB1	0.549

Table A.1 (continued)

ENSG00000221869	CEBPD	0.546
ENSG00000011304	PTBP1	0.546
ENSG00000162733	DDR2	0.545
ENSG00000124570	SERPINB6	0.543
ENSG00000018408	WWTR1	0.543
ENSG00000114698	PLSCR4	0.542
ENSG00000155324	GRAMD3	0.542
ENSG00000136436	CALCOCO2	0.541
ENSG00000185650	ZFP36L1	0.54
ENSG00000069974	RAB27A	0.54
ENSG00000128585	MKLN1	0.539
ENSG00000176783	RUFY1	0.538
ENSG00000068697	LAPTM4A	0.537
ENSG00000150907	FOXO1	0.536
ENSG00000056972	TRAF3IP2	0.535
ENSG00000123096	SSPN	0.535
ENSG00000166483	WEE1	0.534
ENSG00000092531	SNAP23	0.534
ENSG00000103710	RASL12	0.533
ENSG00000147027	TMEM47	0.531
ENSG00000132470	ITGB4	0.531
ENSG00000138434	SSFA2	0.528
ENSG00000182541	LIMK2	0.527
ENSG00000179604	CDC42EP4	0.527
ENSG00000178209	PLEC	0.526
ENSG00000089159	PXN	0.526
ENSG00000110651	CD81	0.524
ENSG00000137193	PIM1	0.523
ENSG00000105854	PON2	0.523
ENSG00000187091	PLCD1	0.523
ENSG00000100906	NFKBIA	0.523
ENSG00000122786	CALD1	0.522
ENSG00000241839	PLEKHO2	0.522
ENSG00000092820	EZR	0.521
ENSG00000125753	VASP	0.519
ENSG00000115325	DOK1	0.518
ENSG00000172201	ID4	0.518
ENSG00000117592	PRDX6	0.518

Table A.1 (continued)

ENSG00000165029	ABCA1	0.518
ENSG00000083168	KAT6A	0.516
ENSG00000152518	ZFP36L2	0.516
ENSG00000162413	KLHL21	0.516
ENSG00000140497	SCAMP2	0.516
ENSG00000100441	KHNYN	0.515
ENSG00000165175	MID1IP1	0.515
ENSG00000163132	MSX1	0.514
ENSG00000170876	TMEM43	0.514
ENSG00000131446	MGAT1	0.513
ENSG00000132669	RIN2	0.513
ENSG00000132329	RAMP1	0.513
ENSG00000152558	TMEM123	0.511
ENSG00000197256	KANK2	0.51
ENSG00000124782	RREB1	0.51
ENSG00000198604	BAZ1A	0.51
ENSG00000163346	PBXIP1	0.509
ENSG00000182718	ANXA2	0.509
ENSG00000011243	AKAP8L	0.509
ENSG00000007372	PAX6	0.507
ENSG00000154188	ANGPT1	0.506
ENSG00000168610	STAT3	0.505
ENSG00000115594	IL1R1	0.505
ENSG00000082781	ITGB5	0.504
ENSG00000177595	PIDD1	0.503
ENSG00000182149	IST1	0.503
ENSG00000061273	HDAC7	0.502
ENSG00000185432	METTL7A	0.502
ENSG00000116044	NFE2L2	0.502
ENSG00000169604	ANTXR1	0.501
ENSG00000070404	FSTL3	0.501
ENSG00000182253	SYNM	0.5
ENSG00000116478	HDAC1	0.498
ENSG00000173991	TCAP	0.497
ENSG00000165458	INPPL1	0.496
ENSG00000049323	LTBP1	0.496
ENSG00000131459	GFPT2	0.495
ENSG00000147065	MSN	0.495

Table A.1 (continued)

ENSG00000141469	SLC14A1	0.494
ENSG00000140575	IQGAP1	0.493
ENSG00000143815	LBR	0.493
ENSG00000118482	PHF3	0.493
ENSG00000136451	VEZF1	0.492
ENSG00000126777	KTN1	0.492
ENSG00000183943	PRKX	0.492
ENSG00000154803	FLCN	0.492
ENSG00000100811	YY1	0.491
ENSG00000115468	EFHD1	0.49
ENSG00000092969	TGFB2	0.49
ENSG00000071967	CYBRD1	0.489
ENSG00000197442	MAP3K5	0.489
ENSG00000182158	CREB3L2	0.488
ENSG00000067141	NEO1	0.488
ENSG00000153914	SREK1	0.488
ENSG00000113732	ATP6V0E1	0.487
ENSG00000157191	NECAP2	0.487
ENSG00000142227	EMP3	0.487
ENSG00000006831	ADIPOR2	0.487
ENSG00000177575	CD163	0.486
ENSG00000198960	ARMCX6	0.485
ENSG00000151491	EPS8	0.484
ENSG00000007384	RHBDF1	0.483
ENSG00000149489	ROM1	0.483
ENSG00000170638	TRABD	0.483
ENSG00000113140	SPARC	0.482
ENSG00000132256	TRIM5	0.482
ENSG00000155368	DBI	0.481
ENSG00000115107	STEAP3	0.481
ENSG00000143819	EPHX1	0.48
ENSG00000173473	SMARCC1	0.479
ENSG00000091409	ITGA6	0.479
ENSG00000133131	MORC4	0.479
ENSG00000127603	MACF1	0.477
ENSG00000168056	LTBP3	0.477
ENSG00000104324	CPQ	0.477
ENSG00000005893	LAMP2	0.476

Table A.1 (continued)

ENSG00000100242	SUN2	0.476
ENSG00000134294	SLC38A2	0.476
ENSG00000140750	ARHGAP17	0.476
ENSG00000064961	HMG20B	0.475
ENSG00000113594	LIFR	0.475
ENSG00000132424	PNISR	0.474
ENSG00000182326	C1S	0.473
ENSG00000163083	INHBB	0.473
ENSG00000073712	FERMT2	0.472
ENSG00000198917	SPOUT1	0.472
ENSG00000154175	ABI3BP	0.472
ENSG00000167191	GPRC5B	0.472
ENSG00000092621	PHGDH	0.47
ENSG00000105281	SLC1A5	0.47
ENSG00000182492	BGN	0.47
ENSG00000185591	SP1	0.47
ENSG00000100014	SPECC1L	0.47
ENSG00000163110	PDLIM5	0.468
ENSG00000128602	SMO	0.468
ENSG00000047457	CP	0.467
ENSG00000173905	GOLIM4	0.466
ENSG00000146648	EGFR	0.466
ENSG00000009830	POMT2	0.466
ENSG00000102125	TAZ	0.465
ENSG00000168994	PXDC1	0.465
ENSG00000135744	AGT	0.465
ENSG00000065883	CDK13	0.464
ENSG00000113658	SMAD5	0.463
ENSG00000005243	COPZ2	0.462
ENSG00000033030	ZCCHC8	0.462
ENSG00000107779	BMPR1A	0.461
ENSG00000079308	TNS1	0.461
ENSG00000178252	WDR6	0.46
ENSG00000171456	ASXL1	0.46
ENSG00000145012	LPP	0.459
ENSG00000131051	RBM39	0.459
ENSG00000148700	ADD3	0.458
ENSG00000000003	TSPAN6	0.457

Table A.1 (continued)

ENSG00000133639	BTG1	0.457
ENSG00000135365	PHF21A	0.457
ENSG00000067955	CBFB	0.457
ENSG00000168077	SCARA3	0.457
ENSG00000139644	TMBIM6	0.456
ENSG00000055070	SZRD1	0.456
ENSG00000079335	CDC14A	0.456
ENSG00000129667	RHBDF2	0.456
ENSG00000104365	IKBKB	0.455
ENSG00000171940	ZNF217	0.455
ENSG00000140299	BNIP2	0.454
ENSG00000006327	TNFRSF12A	0.453
ENSG00000102699	PARP4	0.453
ENSG00000143514	TP53BP2	0.452
ENSG00000163884	KLF15	0.452
ENSG00000148737	TCF7L2	0.451
ENSG00000138119	MYOF	0.451
ENSG00000172037	LAMB2	0.451
ENSG00000087077	TRIP6	0.449
ENSG00000104881	PPP1R13L	0.449
ENSG00000164111	ANXA5	0.448
ENSG00000137145	DENND4C	0.448
ENSG00000112773	FAM46A	0.448
ENSG00000174718	KIAA1551	0.447
ENSG00000172530	BANP	0.447
ENSG00000101680	LAMA1	0.447
ENSG00000021300	PLEKHB1	0.445
ENSG00000205213	LGR4	0.445
ENSG00000011198	ABHD5	0.444
ENSG00000105939	ZC3HAV1	0.443
ENSG00000103449	SALL1	0.443
ENSG00000111321	LTBR	0.443
ENSG00000106538	RARRES2	0.443
ENSG00000171100	MTM1	0.442
ENSG00000163565	IFI16	0.442
ENSG00000172943	PHF8	0.441
ENSG00000104419	NDRG1	0.441
ENSG00000169504	CLIC4	0.441

Table A.1 (continued)

ENSG00000178878	APOLD1	0.441
ENSG00000120690	ELF1	0.441
ENSG00000124749	COL21A1	0.44
ENSG00000097007	ABL1	0.439
ENSG00000156304	SCAF4	0.439
ENSG00000135111	TBX3	0.439
ENSG00000145555	MYO10	0.439
ENSG00000132274	TRIM22	0.439
ENSG00000122863	CHST3	0.438
ENSG00000113300	CNOT6	0.437
ENSG00000008282	SYPL1	0.437
ENSG00000178764	ZHX2	0.437
ENSG00000170370	EMX2	0.436
ENSG00000143418	CERS2	0.436
ENSG00000139218	SCAF11	0.436
ENSG00000079134	THOC1	0.435
ENSG00000054598	FOXC1	0.435
ENSG00000160789	LMNA	0.434
ENSG00000161638	ITGA5	0.434
ENSG00000205403	CFI	0.433
ENSG00000168884	TNIP2	0.433
ENSG00000057252	SOAT1	0.433
ENSG00000154734	ADAMTS1	0.433
ENSG00000129151	BBOX1	0.433
ENSG00000087903	RFX2	0.433
ENSG00000168906	MAT2A	0.432
ENSG00000100813	ACIN1	0.431
ENSG00000129250	KIF1C	0.431
ENSG00000132613	MTSS1L	0.431
ENSG00000106404	CLDN15	0.43
ENSG00000166833	NAV2	0.43
ENSG00000123374	CDK2	0.43
ENSG00000186350	RXRA	0.43
ENSG00000090530	P3H2	0.429
ENSG00000140961	OSGIN1	0.428
ENSG00000135837	CEP350	0.428
ENSG00000170348	TMED10	0.427
ENSG00000100599	RIN3	0.427

Table A.1 (continued)

ENSG00000116747	TROVE2	0.426
ENSG00000104375	STK3	0.426
ENSG00000102908	NFAT5	0.426
ENSG00000106565	TMEM176B	0.426
ENSG00000134574	DDB2	0.425
ENSG00000160862	AZGP1	0.425
ENSG00000149131	SERPING1	0.424
ENSG00000171206	TRIM8	0.424
ENSG00000117360	PRPF3	0.423
ENSG00000138029	HADHB	0.422
ENSG00000102710	SUPT20H	0.422
ENSG00000099860	GADD45B	0.422
ENSG00000114439	BBX	0.422
ENSG00000028137	TNFRSF1B	0.421
ENSG00000141736	ERBB2	0.421
ENSG00000140262	TCF12	0.421
ENSG00000110324	IL10RA	0.419
ENSG00000114353	GNAI2	0.419
ENSG00000148411	NACC2	0.419
ENSG00000143416	SELENBP1	0.417
ENSG00000106333	PCOLCE	0.417
ENSG00000166743	ACSM1	0.417
ENSG00000135929	CYP27A1	0.417
ENSG00000168710	AHCYL1	0.416
ENSG00000163820	FYCO1	0.416
ENSG00000119688	ABCD4	0.416
ENSG00000130309	COLGALT1	0.415
ENSG00000010327	STAB1	0.415
ENSG00000111642	CHD4	0.415
ENSG00000070214	SLC44A1	0.415
ENSG00000124214	STAU1	0.415
ENSG00000134802	SLC43A3	0.414
ENSG00000105355	PLIN3	0.414
ENSG00000156639	ZFAND3	0.413
ENSG00000061936	SFSWAP	0.413
ENSG00000074047	GLI2	0.413
ENSG00000181722	ZBTB20	0.412
ENSG00000149658	YTHDF1	0.412

Table A.1 (continued)

ENSG00000100234	TIMP3	0.412
ENSG00000073614	KDM5A	0.411
ENSG00000165494	PCF11	0.411
ENSG00000139832	RAB20	0.411
ENSG00000157110	RBPMS	0.411
ENSG00000160271	RALGDS	0.411
ENSG00000196924	FLNA	0.411
ENSG00000108691	CCL2	0.408
ENSG00000185499	MUC1	0.408
ENSG00000117280	RAB29	0.408
ENSG00000129219	PLD2	0.408
ENSG00000167978	SRRM2	0.407
ENSG00000250722	SELENOP	0.407
ENSG00000089327	FXYD5	0.407
ENSG00000117523	PRRC2C	0.406
ENSG00000159403	C1R	0.405
ENSG00000173889	PHC3	0.404
ENSG00000117298	ECE1	0.404
ENSG00000100697	DICER1	0.403
ENSG00000076716	GPC4	0.403
ENSG00000138696	BMPR1B	0.403
ENSG00000133321	RARRES3	0.401
ENSG00000177469	PTRF	0.401
ENSG00000168938	PPIC	0.401
ENSG00000118762	PKD2	0.401
ENSG00000161011	SQSTM1	0.4
ENSG00000056998	GYG2	0.4
ENSG00000135473	PAN2	0.4
ENSG00000006747	SCIN	0.4
ENSG00000130147	SH3BP4	0.399
ENSG00000140836	ZFHX3	0.399
ENSG00000184481	FOXO4	0.399
ENSG00000160200	CBS	0.397
ENSG00000101109	STK4	0.397
ENSG00000075234	TTC38	0.396
ENSG00000154380	ENAH	0.396
ENSG00000136205	TNS3	0.396
ENSG00000083312	TNPO1	0.395

Table A.1 (continued)

ENSG00000047644	WWC3	0.395
ENSG00000168899	VAMP5	0.393
ENSG00000078246	TULP3	0.393
ENSG00000188783	PRELP	0.393
ENSG00000145685	LHFPL2	0.393
ENSG00000129654	FOXJ1	0.392
ENSG00000132155	RAF1	0.392
ENSG00000126458	RRAS	0.392
ENSG00000110719	TCIRG1	0.391
ENSG00000106351	AGFG2	0.389
ENSG00000144579	CTDSP1	0.389
ENSG00000135686	KLHL36	0.389
ENSG00000084207	GSTP1	0.388
ENSG00000146425	DYNLT1	0.387
ENSG00000166801	FAM111A	0.387
ENSG00000168003	SLC3A2	0.387
ENSG00000183580	FBXL7	0.385
ENSG00000165792	METTL17	0.385
ENSG00000152049	KCNE4	0.384
ENSG00000109906	ZBTB16	0.382
ENSG00000132561	MATN2	0.382
ENSG00000135862	LAMC1	0.382
ENSG00000187554	TLR5	0.382
ENSG00000134851	TMEM165	0.382
ENSG00000197724	PHF2	0.382
ENSG00000164190	NIPBL	0.381
ENSG00000122862	SRGN	0.381
ENSG00000120693	SMAD9	0.38
ENSG00000152284	TCF7L1	0.38
ENSG00000197405	C5AR1	0.38
ENSG00000127241	MASP1	0.378
ENSG00000147421	HMBOX1	0.378
ENSG00000126070	AGO3	0.378
ENSG00000090382	LYZ	0.378
ENSG00000125952	MAX	0.378
ENSG00000124762	CDKN1A	0.377
ENSG00000114857	NKTR	0.376
ENSG00000120885	CLU	0.376

Table A.1 (continued)

ENSG00000087086	FTL	0.376
ENSG00000137269	LRRC1	0.376
ENSG00000113721	PDGFRB	0.376
ENSG00000112561	TFEB	0.376
ENSG00000101017	CD40	0.375
ENSG00000018280	SLC11A1	0.374
ENSG00000133884	DPF2	0.374
ENSG00000173801	JUP	0.373
ENSG00000087206	UIMC1	0.373
ENSG00000100427	MLC1	0.373
ENSG00000078061	ARAF	0.371
ENSG00000111666	CHPT1	0.37
ENSG00000111450	STX2	0.369
ENSG00000101367	MAPRE1	0.369
ENSG00000065978	YBX1	0.369
ENSG00000063127	SLC6A16	0.368
ENSG00000134744	ZCCHC11	0.368
ENSG00000141519	CCDC40	0.368
ENSG00000171766	GATM	0.366
ENSG00000169249	ZRSR2	0.366
ENSG00000139842	CUL4A	0.365
ENSG00000060237	WNK1	0.365
ENSG00000167601	AXL	0.365
ENSG00000136938	ANP32B	0.364
ENSG00000106397	PLOD3	0.364
ENSG00000134815	DHX34	0.364
ENSG00000172936	MYD88	0.363
ENSG00000124145	SDC4	0.362
ENSG00000160712	IL6R	0.362
ENSG00000089472	HEPH	0.362
ENSG00000165959	CLMN	0.362
ENSG00000182185	RAD51B	0.362
ENSG00000172830	SSH3	0.361
ENSG00000097033	SH3GLB1	0.361
ENSG00000136997	MYC	0.36
ENSG00000128604	IRF5	0.36
ENSG00000142173	COL6A2	0.36
ENSG00000130821	SLC6A8	0.36

Table A.1 (continued)

ENSG00000131748	STARD3	0.359
ENSG00000130055	GDPD2	0.359
ENSG00000011600	TYROBP	0.357
ENSG00000177051	FBXO46	0.357
ENSG00000186204	CYP4F12	0.357
ENSG00000013364	MVP	0.357
ENSG00000126803	HSPA2	0.357
ENSG00000151322	NPAS3	0.357
ENSG00000108773	KAT2A	0.356
ENSG00000070778	PTPN21	0.356
ENSG00000158186	MRAS	0.355
ENSG00000167491	GATAD2A	0.355
ENSG00000020633	RUNX3	0.352
ENSG00000166333	ILK	0.352
ENSG00000159140	SON	0.352
ENSG00000167994	RAB3IL1	0.352
ENSG00000113269	RNF130	0.35
ENSG00000138080	EMILIN1	0.35
ENSG00000154217	PITPNC1	0.348
ENSG00000155926	SLA	0.348
ENSG00000169403	PTAFR	0.347
ENSG00000084754	HADHA	0.347
ENSG00000100852	ARHGAP5	0.347
ENSG00000103202	NME4	0.346
ENSG00000100393	EP300	0.345
ENSG00000120896	SORBS3	0.345
ENSG00000128591	FLNC	0.345
ENSG00000128274	A4GALT	0.345
ENSG00000168476	REEP4	0.345
ENSG00000083857	FAT1	0.344
ENSG00000042493	CAPG	0.344
ENSG00000171700	RGS19	0.344
ENSG00000125347	IRF1	0.343
ENSG00000136732	GYPC	0.343
ENSG00000108846	ABCC3	0.342
ENSG00000120594	PLXDC2	0.341
ENSG00000077238	IL4R	0.341
ENSG00000066468	FGFR2	0.341

Table A.1 (continued)

ENSG00000165025	SYK	0.34
ENSG00000184113	CLDN5	0.34
ENSG00000133488	SEC14L4	0.339
ENSG00000141458	NPC1	0.339
ENSG00000118557	PMFBP1	0.339
ENSG00000102265	TIMP1	0.339
ENSG00000144677	CTDSPL	0.338
ENSG00000076685	NT5C2	0.338
ENSG00000086288	NME8	0.338
ENSG00000100918	REC8	0.338
ENSG00000167995	BEST1	0.337
ENSG00000141510	TP53	0.337
ENSG00000165424	ZCCHC24	0.337
ENSG00000162231	NXF1	0.337
ENSG00000114904	NEK4	0.335
ENSG00000008294	SPAG9	0.335
ENSG00000132024	CC2D1A	0.334
ENSG00000105137	SYDE1	0.334
ENSG00000072121	ZFYVE26	0.334
ENSG00000154240	CEP112	0.333
ENSG00000184014	DENND5A	0.333
ENSG00000184557	SOCS3	0.333
ENSG00000100227	POLDIP3	0.333
ENSG00000133574	GIMAP4	0.332
ENSG00000106100	NOD1	0.331
ENSG00000114331	ACAP2	0.331
ENSG00000130669	PAK4	0.329
ENSG00000114796	KLHL24	0.329
ENSG00000107968	MAP3K8	0.329
ENSG00000148400	NOTCH1	0.328
ENSG00000012983	MAP4K5	0.327
ENSG00000198088	NUP62CL	0.327
ENSG00000151748	SAV1	0.326
ENSG00000180447	GAS1	0.324
ENSG00000136861	CDK5RAP2	0.324
ENSG00000102359	SRPX2	0.324
ENSG00000130164	LDLR	0.324
ENSG00000188269	OR7A5	0.323

Table A.1 (continued)

ENSG00000134548	SPX	0.322
ENSG00000126246	IGFLR1	0.321
ENSG00000064393	HIPK2	0.319
ENSG00000180353	HCLS1	0.317
ENSG00000005339	CREBBP	0.317
ENSG00000166387	PPFIBP2	0.316
ENSG00000010292	NCAPD2	0.316
ENSG00000090554	FLT3LG	0.316
ENSG00000188404	SELL	0.315
ENSG00000107731	UNC5B	0.315
ENSG00000123066	MED13L	0.315
ENSG00000141934	PLPP2	0.314
ENSG00000127483	HP1BP3	0.314
ENSG00000112936	C7	0.312
ENSG00000114315	HES1	0.312
ENSG00000174799	CEP135	0.311
ENSG00000162407	PLPP3	0.311
ENSG00000172164	SNTB1	0.31
ENSG00000105698	USF2	0.309
ENSG00000132825	PPP1R3D	0.308
ENSG00000072778	ACADVL	0.308
ENSG00000141756	FKBP10	0.308
ENSG00000104894	CD37	0.308
ENSG00000113328	CCNG1	0.306
ENSG00000071626	DAZAP1	0.304
ENSG00000155465	SLC7A7	0.303
ENSG00000164776	PHKG1	0.303
ENSG00000115085	ZAP70	0.303
ENSG00000138376	BARD1	0.302
ENSG00000120733	KDM3B	0.301
ENSG00000166224	SGPL1	0.301
ENSG00000105372	RPS19	0.3
ENSG00000134222	PSRC1	0.3
ENSG00000131669	NINJ1	0.3
ENSG00000107902	LHPP	0.299
ENSG00000185043	CIB1	0.298
ENSG00000173805	HAP1	0.296
ENSG00000115648	MLPH	0.296

Table A.1 (continued)

ENSG00000105229	PIAS4	0.296
ENSG00000164199	ADGRV1	0.295
ENSG00000110057	UNC93B1	0.295
ENSG00000173068	BNC2	0.293
ENSG00000118503	TNFAIP3	0.292
ENSG00000118523	CTGF	0.292
ENSG00000146859	TMEM140	0.292
ENSG00000128284	APOL3	0.288
ENSG00000136826	KLF4	0.288
ENSG00000175826	CTDNEP1	0.286
ENSG00000099998	GGT5	0.286
ENSG00000126016	AMOT	0.284
ENSG00000167766	ZNF83	0.284
ENSG00000146247	PHIP	0.284
ENSG00000012223	LTF	0.283
ENSG00000072694	FCGR2B	0.281
ENSG00000049540	ELN	0.28
ENSG00000101605	MYOM1	0.28
ENSG00000151726	ACSL1	0.278
ENSG00000172340	SUCLG2	0.276
ENSG00000131634	TMEM204	0.274
ENSG00000105366	SIGLEC8	0.274
ENSG00000143344	RGL1	0.273
ENSG00000114166	KAT2B	0.271
ENSG00000157637	SLC38A10	0.271
ENSG00000163694	RBM47	0.27
ENSG00000156150	ALX3	0.269
ENSG00000101336	HCK	0.268
ENSG00000150048	CLEC1A	0.267
ENSG00000121361	KCNJ8	0.267
ENSG00000138379	MSTN	0.267
ENSG00000125730	C3	0.265
ENSG00000171606	ZNF274	0.262
ENSG00000031823	RANBP3	0.261
ENSG00000021826	CPS1	0.257
ENSG00000120458	MSANTD2	0.256
ENSG00000040531	CTNS	0.256
ENSG00000177084	POLE	0.253

Table A.1 (continued)

ENSG00000176485	PLA2G16	0.253
ENSG00000175591	P2RY2	0.248
ENSG00000154553	PDLIM3	0.248
ENSG00000091527	CDV3	0.247
ENSG00000081377	CDC14B	0.247
ENSG00000021355	SERPINB1	0.243
ENSG00000168685	IL7R	0.242
ENSG00000105383	CD33	0.242
ENSG00000068079	IFI35	0.237
ENSG00000111837	MAK	0.234
ENSG00000170891	CYTL1	0.233
ENSG00000016602	CLCA4	0.23
ENSG00000166888	STAT6	0.229
ENSG00000099139	PCSK5	0.228
ENSG00000108405	P2RX1	0.227
ENSG00000175793	SFN	0.222
ENSG00000171962	DRC3	0.217
ENSG00000135547	HEY2	0.217
ENSG00000033327	GAB2	0.214
ENSG00000082996	RNF13	0.209
ENSG00000197093	GAL3ST4	0.205
ENSG00000011422	PLAUR	0.204
ENSG00000139292	LGR5	0.204
ENSG00000198000	NOL8	0.198
ENSG00000147257	GPC3	0.197
ENSG00000145779	TNFAIP8	0.196
ENSG00000010671	BTK	0.195
ENSG00000100784	RPS6KA5	0.192
ENSG00000108848	LUC7L3	0.192
ENSG00000132702	HAPLN2	0.187
ENSG00000078674	PCM1	0.186
ENSG00000110079	MS4A4A	0.182
ENSG00000180644	PRF1	0.171
ENSG00000074356	NCBP3	0.17
ENSG00000070759	TESK2	0.167
ENSG00000135720	DYNC1LI2	0.163
ENSG00000137731	FXYD2	0.16
ENSG00000102580	DNAJC3	0.147

Table A.1 (continued)

ENSG00000106012	IQCE	0.088
ENSG00000142864	SERBP1	0.08
ENSG00000161202	DVL3	0.072

APPENDIX B

LIST OF AD-RELATED GENES SHOWING DOWNREGULATION TREND

Table B.1: List of genes showing significant dowregulation trend across datasets. *M-rho* is the median rho value of a gene among datasets.

ENSG ID	**Gene Name**	***M-rho***
ENSG00000124785	NRN1	-0.663
ENSG00000175426	PCSK1	-0.656
ENSG00000144834	TAGLN3	-0.641
ENSG00000135119	RNFT2	-0.631
ENSG00000131100	ATP6V1E1	-0.63
ENSG00000111652	COPS7A	-0.629
ENSG00000100983	GSS	-0.623
ENSG00000110148	CCKBR	-0.619
ENSG00000198932	GPRASP1	-0.617
ENSG00000032389	TSSC1	-0.613
ENSG00000165704	HPRT1	-0.602
ENSG00000149269	PAK1	-0.601
ENSG00000089123	TASP1	-0.6
ENSG00000117152	RGS4	-0.599
ENSG00000156395	SORCS3	-0.598
ENSG00000021645	NRXN3	-0.597
ENSG00000164600	NEUROD6	-0.597
ENSG00000089199	CHGB	-0.594
ENSG00000053372	MRTO4	-0.591
ENSG00000086717	PPEF1	-0.591

Table B.1 (continued)

ENSG00000232859	LYRM9	-0.589
ENSG00000163032	VSNL1	-0.589
ENSG00000068615	REEP1	-0.587
ENSG00000003147	ICA1	-0.587
ENSG00000132434	LANCL2	-0.582
ENSG00000159720	ATP6V0D1	-0.582
ENSG00000175352	NRIP3	-0.58
ENSG00000162975	KCNF1	-0.578
ENSG00000181991	MRPS11	-0.578
ENSG00000114279	FGF12	-0.577
ENSG00000136261	BZW2	-0.576
ENSG00000185518	SV2B	-0.576
ENSG00000156298	TSPAN7	-0.575
ENSG00000187601	MAGEH1	-0.575
ENSG00000117069	ST6GALNAC5	-0.575
ENSG00000174938	SEZ6L2	-0.574
ENSG00000105171	POP4	-0.574
ENSG00000082458	DLG3	-0.572
ENSG00000162188	GNG3	-0.572
ENSG00000125166	GOT2	-0.57
ENSG00000133318	RTN3	-0.569
ENSG00000148481	FAM188A	-0.569
ENSG00000171617	ENC1	-0.568
ENSG00000198836	OPA1	-0.567
ENSG00000079841	RIMS1	-0.565
ENSG00000136827	TOR1A	-0.563
ENSG00000143198	MGST3	-0.561
ENSG00000103316	CRYM	-0.561
ENSG00000111669	TPI1	-0.559
ENSG00000116141	MARK1	-0.558
ENSG00000119866	BCL11A	-0.556
ENSG00000172348	RCAN2	-0.556
ENSG00000102312	PORCN	-0.556
ENSG00000164885	CDK5	-0.556
ENSG00000091972	CD200	-0.555
ENSG00000213626	LBH	-0.555
ENSG00000132305	IMMT	-0.555

Table B.1 (continued)

ENSG00000145708	CRHBP	-0.553
ENSG00000133083	DCLK1	-0.552
ENSG00000100554	ATP6V1D	-0.552
ENSG00000104435	STMN2	-0.552
ENSG00000078328	RBFOX1	-0.551
ENSG00000136928	GABBR2	-0.551
ENSG00000105409	ATP1A3	-0.55
ENSG00000103056	SMPD3	-0.55
ENSG00000173726	TOMM20	-0.55
ENSG00000137815	RTF1	-0.55
ENSG00000050748	MAPK9	-0.549
ENSG00000198954	KIF1BP	-0.548
ENSG00000173465	SSSCA1	-0.546
ENSG00000179222	MAGED1	-0.546
ENSG00000086300	SNX10	-0.545
ENSG00000165916	PSMC3	-0.545
ENSG00000006116	CACNG3	-0.545
ENSG00000157895	C12orf43	-0.544
ENSG00000135750	KCNK1	-0.544
ENSG00000010256	UQCRC1	-0.544
ENSG00000157152	SYN2	-0.543
ENSG00000139180	NDUFA9	-0.543
ENSG00000022355	GABRA1	-0.542
ENSG00000169599	NFU1	-0.54
ENSG00000174446	SNAPC5	-0.538
ENSG00000123178	SPRYD7	-0.538
ENSG00000013016	EHD3	-0.537
ENSG00000136521	NDUFB5	-0.537
ENSG00000176697	BDNF	-0.537
ENSG00000123297	TSFM	-0.536
ENSG00000132932	ATP8A2	-0.536
ENSG00000163875	MEAF6	-0.536
ENSG00000104738	MCM4	-0.534
ENSG00000100216	TOMM22	-0.534
ENSG00000172115	CYCS	-0.533
ENSG00000014824	SLC30A9	-0.533
ENSG00000079785	DDX1	-0.532
ENSG00000165152	TMEM246	-0.532

Table B.1 (continued)

ENSG00000159199	ATP5G1	-0.532
ENSG00000187678	SPRY4	-0.531
ENSG00000147852	VLDLR	-0.53
ENSG00000127445	PIN1	-0.53
ENSG00000170899	GSTA4	-0.529
ENSG00000213190	MLLT11	-0.528
ENSG00000099341	PSMD8	-0.528
ENSG00000129636	ITFG1	-0.528
ENSG00000113327	GABRG2	-0.527
ENSG00000130638	ATXN10	-0.526
ENSG00000091483	FH	-0.525
ENSG00000153233	PTPRR	-0.524
ENSG00000172575	RASGRP1	-0.524
ENSG00000126970	ZC4H2	-0.523
ENSG00000198794	SCAMP5	-0.523
ENSG00000133627	ACTR3B	-0.523
ENSG00000040341	STAU2	-0.522
ENSG00000143499	SMYD2	-0.521
ENSG00000148798	INA	-0.521
ENSG00000124659	TBCC	-0.521
ENSG00000172336	POP7	-0.52
ENSG00000088812	ATRN	-0.519
ENSG00000125354	SEPT6	-0.518
ENSG00000169139	UBE2V2	-0.518
ENSG00000112186	CAP2	-0.517
ENSG00000143158	MPC2	-0.517
ENSG00000171951	SCG2	-0.517
ENSG00000204856	FAM216A	-0.516
ENSG00000113068	PFDN1	-0.515
ENSG00000120053	GOT1	-0.515
ENSG00000107518	ATRNL1	-0.515
ENSG00000161203	AP2M1	-0.515
ENSG00000174684	B4GAT1	-0.515
ENSG00000004779	NDUFAB1	-0.515
ENSG00000156253	RWDD2B	-0.514
ENSG00000149809	TM7SF2	-0.514
ENSG00000118971	CCND2	-0.514
ENSG00000091640	SPAG7	-0.514

Table B.1 (continued)

ENSG00000152495	CAMK4	-0.514
ENSG00000108384	RAD51C	-0.514
ENSG00000102078	SLC25A14	-0.513
ENSG00000132639	SNAP25	-0.513
ENSG00000158560	DYNC1I1	-0.512
ENSG00000143106	PSMA5	-0.512
ENSG00000139637	C12orf10	-0.512
ENSG00000127252	HRASLS	-0.511
ENSG00000128683	GAD1	-0.511
ENSG00000125962	ARMCX5	-0.511
ENSG00000104643	MTMR9	-0.511
ENSG00000168546	GFRA2	-0.51
ENSG00000082213	C5orf22	-0.51
ENSG00000153558	FBXL2	-0.509
ENSG00000180543	TSPYL5	-0.509
ENSG00000196876	SCN8A	-0.508
ENSG00000105393	BABAM1	-0.508
ENSG00000102003	SYP	-0.508
ENSG00000010818	HIVEP2	-0.506
ENSG00000118276	B4GALT6	-0.505
ENSG00000165629	ATP5C1	-0.504
ENSG00000154723	ATP5J	-0.504
ENSG00000181929	PRKAG1	-0.504
ENSG00000131473	ACLY	-0.503
ENSG00000175906	ARL4D	-0.503
ENSG00000109158	GABRA4	-0.503
ENSG00000103723	AP3B2	-0.502
ENSG00000145730	PAM	-0.501
ENSG00000168032	ENTPD3	-0.501
ENSG00000136750	GAD2	-0.5
ENSG00000061918	GUCY1B3	-0.5
ENSG00000101654	RNMT	-0.5
ENSG00000188690	UROS	-0.5
ENSG00000157064	NMNAT2	-0.499
ENSG00000171703	TCEA2	-0.499
ENSG00000166257	SCN3B	-0.499
ENSG00000130540	SULT4A1	-0.498
ENSG00000138663	COPS4	-0.498

Table B.1 (continued)

ENSG00000138757	G3BP2	-0.498
ENSG00000107105	ELAVL2	-0.497
ENSG00000006128	TAC1	-0.496
ENSG00000189159	HN1	-0.496
ENSG00000139405	RITA1	-0.495
ENSG00000085415	SEH1L	-0.493
ENSG00000109738	GLRB	-0.493
ENSG00000136950	ARPC5L	-0.492
ENSG00000085377	PREP	-0.492
ENSG00000159082	SYNJ1	-0.492
ENSG00000130558	OLFM1	-0.491
ENSG00000125863	MKKS	-0.491
ENSG00000114405	C3orf14	-0.491
ENSG00000175182	FAM131A	-0.491
ENSG00000145632	PLK2	-0.489
ENSG00000182220	ATP6AP2	-0.489
ENSG00000163399	ATP1A1	-0.489
ENSG00000114948	ADAM23	-0.489
ENSG00000103769	RAB11A	-0.488
ENSG00000214517	PPME1	-0.488
ENSG00000089818	NECAP1	-0.488
ENSG00000013392	RWDD2A	-0.487
ENSG00000147416	ATP6V1B2	-0.487
ENSG00000167515	TRAPPC2L	-0.487
ENSG00000116106	EPHA4	-0.486
ENSG00000105223	PLD3	-0.486
ENSG00000112695	COX7A2	-0.486
ENSG00000066382	MPPED2	-0.486
ENSG00000156411	C14orf2	-0.486
ENSG00000127561	SYNGR3	-0.486
ENSG00000110429	FBXO3	-0.485
ENSG00000104112	SCG3	-0.484
ENSG00000175175	PPM1E	-0.483
ENSG00000060982	BCAT1	-0.483
ENSG00000183520	UTP11	-0.483
ENSG00000112367	FIG4	-0.483
ENSG00000152556	PFKM	-0.483
ENSG00000114021	NIT2	-0.483

Table B.1 (continued)

ENSG00000136854	STXBP1	-0.483
ENSG00000065609	SNAP91	-0.482
ENSG00000134809	TIMM10	-0.482
ENSG00000105696	TMEM59L	-0.482
ENSG00000104093	DMXL2	-0.482
ENSG00000101247	NDUFAF5	-0.481
ENSG00000171303	KCNK3	-0.48
ENSG00000144645	OSBPL10	-0.48
ENSG00000138207	RBP4	-0.479
ENSG00000198961	PJA2	-0.479
ENSG00000174437	ATP2A2	-0.479
ENSG00000152214	RIT2	-0.479
ENSG00000071553	ATP6AP1	-0.479
ENSG00000167863	ATP5H	-0.478
ENSG00000075340	ADD2	-0.478
ENSG00000122873	CISD1	-0.478
ENSG00000131779	PEX11B	-0.478
ENSG00000110931	CAMKK2	-0.477
ENSG00000162377	COA7	-0.477
ENSG00000137547	MRPL15	-0.476
ENSG00000101079	NDRG3	-0.476
ENSG00000134375	TIMM17A	-0.475
ENSG00000152092	ASTN1	-0.475
ENSG00000128654	MTX2	-0.475
ENSG00000060709	RIMBP2	-0.475
ENSG00000151806	GUF1	-0.474
ENSG00000171208	NETO2	-0.474
ENSG00000040933	INPP4A	-0.474
ENSG00000100276	RASL10A	-0.474
ENSG00000168438	CDC40	-0.474
ENSG00000170445	HARS	-0.474
ENSG00000171189	GRIK1	-0.473
ENSG00000077522	ACTN2	-0.473
ENSG00000182013	PNMAL1	-0.472
ENSG00000147669	POLR2K	-0.471
ENSG00000179091	CYC1	-0.471
ENSG00000100897	DCAF11	-0.47
ENSG00000110427	KIAA1549L	-0.47

Table B.1 (continued)

ENSG00000169213	RAB3B	-0.469
ENSG00000137996	RTCA	-0.469
ENSG00000087095	NLK	-0.469
ENSG00000060140	STYK1	-0.468
ENSG00000072832	CRMP1	-0.468
ENSG00000116459	ATP5F1	-0.468
ENSG00000073803	MAP3K13	-0.467
ENSG00000123415	SMUG1	-0.466
ENSG00000164258	NDUFS4	-0.466
ENSG00000133026	MYH10	-0.466
ENSG00000116918	TSNAX	-0.466
ENSG00000176871	WSB2	-0.466
ENSG00000162735	PEX19	-0.465
ENSG00000144635	DYNC1LI1	-0.462
ENSG00000006625	GGCT	-0.461
ENSG00000091157	WDR7	-0.461
ENSG00000107758	PPP3CB	-0.461
ENSG00000055163	CYFIP2	-0.461
ENSG00000006468	ETV1	-0.46
ENSG00000173692	PSMD1	-0.459
ENSG00000166669	ATF7IP2	-0.459
ENSG00000150768	DLAT	-0.459
ENSG00000075945	KIFAP3	-0.457
ENSG00000054356	PTPRN	-0.457
ENSG00000138028	CGREF1	-0.457
ENSG00000220205	VAMP2	-0.456
ENSG00000102226	USP11	-0.455
ENSG00000104381	GDAP1	-0.455
ENSG00000095002	MSH2	-0.454
ENSG00000101638	ST8SIA5	-0.454
ENSG00000143786	CNIH3	-0.453
ENSG00000047249	ATP6V1H	-0.453
ENSG00000120875	DUSP4	-0.452
ENSG00000115828	QPCT	-0.452
ENSG00000140284	SLC27A2	-0.451
ENSG00000145725	PPIP5K2	-0.45
ENSG00000155959	VBP1	-0.45
ENSG00000175602	CCDC85B	-0.45

Table B.1 (continued)

ENSG00000096092	TMEM14A	-0.45
ENSG00000137055	PLAA	-0.448
ENSG00000101266	CSNK2A1	-0.448
ENSG00000140945	CDH13	-0.448
ENSG00000136718	IMP4	-0.448
ENSG00000047597	XK	-0.447
ENSG00000172020	GAP43	-0.446
ENSG00000151690	MFSD6	-0.446
ENSG00000132423	COQ3	-0.446
ENSG00000170290	SLN	-0.446
ENSG00000049759	NEDD4L	-0.446
ENSG00000145242	EPHA5	-0.445
ENSG00000114573	ATP6V1A	-0.445
ENSG00000184672	RALYL	-0.445
ENSG00000153823	PID1	-0.445
ENSG00000127463	EMC1	-0.445
ENSG00000198689	SLC9A6	-0.444
ENSG00000067715	SYT1	-0.443
ENSG00000138311	ZNF365	-0.443
ENSG00000166902	MRPL16	-0.443
ENSG00000114023	FAM162A	-0.443
ENSG00000168291	PDHB	-0.442
ENSG00000108528	SLC25A11	-0.442
ENSG00000184076	UQCR10	-0.442
ENSG00000150787	PTS	-0.442
ENSG00000128656	CHN1	-0.442
ENSG00000008277	ADAM22	-0.441
ENSG00000159259	CHAF1B	-0.441
ENSG00000136463	TACO1	-0.44
ENSG00000154162	CDH12	-0.44
ENSG00000168538	TRAPPC11	-0.44
ENSG00000104888	SLC17A7	-0.439
ENSG00000162374	ELAVL4	-0.439
ENSG00000068366	ACSL4	-0.438
ENSG00000073670	ADAM11	-0.438
ENSG00000076554	TPD52	-0.437
ENSG00000181852	RNF41	-0.436
ENSG00000118402	ELOVL4	-0.434

Table B.1 (continued)

ENSG00000182636	NDN	-0.434
ENSG00000213619	NDUFS3	-0.434
ENSG00000157087	ATP2B2	-0.434
ENSG00000131507	NDFIP1	-0.434
ENSG00000186462	NAP1L2	-0.433
ENSG00000103034	NDRG4	-0.432
ENSG00000141367	CLTC	-0.431
ENSG00000112290	WASF1	-0.431
ENSG00000017427	IGF1	-0.431
ENSG00000183715	OPCML	-0.43
ENSG00000164129	NPY5R	-0.43
ENSG00000145916	RMND5B	-0.43
ENSG00000115840	SLC25A12	-0.429
ENSG00000126214	KLC1	-0.429
ENSG00000112146	FBXO9	-0.427
ENSG00000109670	FBXW7	-0.427
ENSG00000077348	EXOSC5	-0.426
ENSG00000091140	DLD	-0.425
ENSG00000109832	DDX25	-0.425
ENSG00000165678	GHITM	-0.425
ENSG00000145293	ENOPH1	-0.425
ENSG00000133135	RNF128	-0.424
ENSG00000138686	BBS7	-0.424
ENSG00000075089	ACTR6	-0.423
ENSG00000104723	TUSC3	-0.422
ENSG00000004897	CDC27	-0.422
ENSG00000186487	MYT1L	-0.422
ENSG00000163577	EIF5A2	-0.42
ENSG00000157542	KCNJ6	-0.419
ENSG00000141030	COPS3	-0.419
ENSG00000102678	FGF9	-0.418
ENSG00000107242	PIP5K1B	-0.418
ENSG00000163618	CADPS	-0.417
ENSG00000163624	CDS1	-0.417
ENSG00000106341	PPP1R17	-0.416
ENSG00000136243	NUPL2	-0.416
ENSG00000198356	ASNA1	-0.416
ENSG00000112237	CCNC	-0.414

Table B.1 (continued)

ENSG00000180875	GREM2	-0.414
ENSG00000078369	GNB1	-0.414
ENSG00000121769	FABP3	-0.413
ENSG00000184408	KCND2	-0.413
ENSG00000128245	YWHAH	-0.412
ENSG00000147571	CRH	-0.412
ENSG00000183036	PCP4	-0.411
ENSG00000087470	DNM1L	-0.411
ENSG00000196482	ESRRG	-0.409
ENSG00000162989	KCNJ3	-0.409
ENSG00000175110	MRPS22	-0.409
ENSG00000108684	ASIC2	-0.408
ENSG00000169255	B3GALNT1	-0.408
ENSG00000101977	MCF2	-0.407
ENSG00000134440	NARS	-0.407
ENSG00000196290	NIF3L1	-0.407
ENSG00000172209	GPR22	-0.405
ENSG00000013561	RNF14	-0.405
ENSG00000112992	NNT	-0.403
ENSG00000114544	SLC41A3	-0.402
ENSG00000109919	MTCH2	-0.402
ENSG00000013503	POLR3B	-0.401
ENSG00000138095	LRPPRC	-0.399
ENSG00000164209	SLC25A46	-0.399
ENSG00000170456	DENND5B	-0.398
ENSG00000155966	AFF2	-0.398
ENSG00000177971	IMP3	-0.398
ENSG00000100095	SEZ6L	-0.398
ENSG00000065154	OAT	-0.397
ENSG00000170522	ELOVL6	-0.397
ENSG00000023330	ALAS1	-0.395
ENSG00000112293	GPLD1	-0.395
ENSG00000085365	SCAMP1	-0.394
ENSG00000118432	CNR1	-0.394
ENSG00000139719	VPS33A	-0.394
ENSG00000164815	ORC5	-0.394
ENSG00000179915	NRXN1	-0.393
ENSG00000139910	NOVA1	-0.392

Table B.1 (continued)

ENSG00000110435	PDHX	-0.392
ENSG00000117245	KIF17	-0.391
ENSG00000091844	RGS17	-0.39
ENSG00000119812	FAM98A	-0.39
ENSG00000121897	LIAS	-0.39
ENSG00000115233	PSMD14	-0.39
ENSG00000139874	SSTR1	-0.389
ENSG00000166197	NOLC1	-0.389
ENSG00000126950	TMEM35A	-0.389
ENSG00000118939	UCHL3	-0.388
ENSG00000104442	ARMC1	-0.386
ENSG00000124140	SLC12A5	-0.386
ENSG00000156515	HK1	-0.384
ENSG00000121871	SLITRK3	-0.383
ENSG00000161204	ABCF3	-0.383
ENSG00000184867	ARMCX2	-0.383
ENSG00000125827	TMX4	-0.382
ENSG00000183665	TRMT12	-0.38
ENSG00000129625	REEP5	-0.38
ENSG00000170231	FABP6	-0.379
ENSG00000123352	SPATS2	-0.379
ENSG00000174405	LIG4	-0.379
ENSG00000162694	EXTL2	-0.377
ENSG00000136045	PWP1	-0.377
ENSG00000100823	APEX1	-0.376
ENSG00000124194	GDAP1L1	-0.376
ENSG00000145681	HAPLN1	-0.376
ENSG00000154277	UCHL1	-0.375
ENSG00000170791	CHCHD7	-0.371
ENSG00000196277	GRM7	-0.371
ENSG00000065665	SEC61A2	-0.37
ENSG00000151500	THYN1	-0.369
ENSG00000090263	MRPS33	-0.369
ENSG00000005249	PRKAR2B	-0.369
ENSG00000168824	NSG1	-0.369
ENSG00000100271	TTLL1	-0.368
ENSG00000105568	PPP2R1A	-0.367
ENSG00000174842	GLMN	-0.366

Table B.1 (continued)

ENSG00000152642	GPD1L	-0.366
ENSG00000119723	COQ6	-0.366
ENSG00000134265	NAPG	-0.366
ENSG00000006740	ARHGAP44	-0.366
ENSG00000114520	SNX4	-0.366
ENSG00000198554	WDHD1	-0.365
ENSG00000123091	RNF11	-0.365
ENSG00000108176	DNAJC12	-0.365
ENSG00000106976	DNM1	-0.364
ENSG00000180720	CHRM4	-0.364
ENSG00000147650	LRP12	-0.364
ENSG00000165495	PKNOX2	-0.364
ENSG00000164163	ABCE1	-0.364
ENSG00000186081	KRT5	-0.362
ENSG00000144136	SLC20A1	-0.362
ENSG00000204262	COL5A2	-0.36
ENSG00000157193	LRP8	-0.36
ENSG00000105778	AVL9	-0.357
ENSG00000109255	NMU	-0.357
ENSG00000113161	HMGCR	-0.353
ENSG00000153310	FAM49B	-0.351
ENSG00000113360	DROSHA	-0.35
ENSG00000151247	EIF4E	-0.349
ENSG00000083750	RRAGB	-0.349
ENSG00000188021	UBQLN2	-0.346
ENSG00000155511	GRIA1	-0.346
ENSG00000136738	STAM	-0.346
ENSG00000178896	EXOSC4	-0.345
ENSG00000198825	INPP5F	-0.345
ENSG00000106013	ANKRD7	-0.344
ENSG00000138069	RAB1A	-0.344
ENSG00000168496	FEN1	-0.343
ENSG00000137274	BPHL	-0.342
ENSG00000102144	PGK1	-0.342
ENSG00000166848	TERF2IP	-0.341
ENSG00000113312	TTC1	-0.341
ENSG00000181789	COPG1	-0.34
ENSG00000090932	DLL3	-0.34

Table B.1 (continued)

ENSG00000104231	ZFAND1	-0.339
ENSG00000133119	RFC3	-0.337
ENSG00000147642	SYBU	-0.336
ENSG00000186310	NAP1L3	-0.335
ENSG00000143774	GUK1	-0.334
ENSG00000165672	PRDX3	-0.331
ENSG00000101132	PFDN4	-0.331
ENSG00000125629	INSIG2	-0.33
ENSG00000169760	NLGN1	-0.33
ENSG00000167862	MRPL58	-0.328
ENSG00000101856	PGRMC1	-0.327
ENSG00000100934	SEC23A	-0.327
ENSG00000135250	SRPK2	-0.326
ENSG00000137252	HCRTR2	-0.326
ENSG00000090266	NDUFB2	-0.325
ENSG00000101746	NOL4	-0.325
ENSG00000101365	IDH3B	-0.324
ENSG00000198369	SPRED2	-0.324
ENSG00000106537	TSPAN13	-0.323
ENSG00000067842	ATP2B3	-0.323
ENSG00000112697	TMEM30A	-0.322
ENSG00000121964	GTDC1	-0.32
ENSG00000092108	SCFD1	-0.32
ENSG00000006210	CX3CL1	-0.319
ENSG00000108389	MTMR4	-0.315
ENSG00000156471	PTDSS1	-0.315
ENSG00000108924	HLF	-0.314
ENSG00000139505	MTMR6	-0.314
ENSG00000097046	CDC7	-0.314
ENSG00000146476	ARMT1	-0.311
ENSG00000066777	ARFGEF1	-0.311
ENSG00000011083	SLC6A7	-0.309
ENSG00000162630	B3GALT2	-0.308
ENSG00000163947	ARHGEF3	-0.307
ENSG00000141404	GNAL	-0.307
ENSG00000047621	C12orf4	-0.307
ENSG00000170633	RNF34	-0.306
ENSG00000113643	RARS	-0.306

Table B.1 (continued)

ENSG00000126243	LRFN3	-0.306
ENSG00000117155	SSX2IP	-0.305
ENSG00000113100	CDH9	-0.304
ENSG00000104863	LIN7B	-0.303
ENSG00000198046	ZNF667	-0.299
ENSG00000164068	RNF123	-0.297
ENSG00000100567	PSMA3	-0.296
ENSG00000050438	SLC4A8	-0.296
ENSG00000164100	NDST3	-0.296
ENSG00000243147	MRPL33	-0.294
ENSG00000115365	LANCL1	-0.294
ENSG00000106683	LIMK1	-0.294
ENSG00000198648	STK39	-0.29
ENSG00000144711	IQSEC1	-0.288
ENSG00000172331	BPGM	-0.288
ENSG00000120437	ACAT2	-0.287
ENSG00000072041	SLC6A15	-0.286
ENSG00000100749	VRK1	-0.285
ENSG00000177733	HNRNPA0	-0.285
ENSG00000151835	SACS	-0.284
ENSG00000134318	ROCK2	-0.284
ENSG00000067829	IDH3G	-0.282
ENSG00000105364	MRPL4	-0.281
ENSG00000082482	KCNK2	-0.281
ENSG00000106355	LSM5	-0.28
ENSG00000180530	NRIP1	-0.278
ENSG00000182134	TDRKH	-0.278
ENSG00000123983	ACSL3	-0.275
ENSG00000126247	CAPNS1	-0.275
ENSG00000141098	GFOD2	-0.275
ENSG00000163541	SUCLG1	-0.273
ENSG00000147124	ZNF41	-0.273
ENSG00000119705	SLIRP	-0.272
ENSG00000198300	PEG3	-0.271
ENSG00000154040	CABYR	-0.269
ENSG00000025772	TOMM34	-0.269
ENSG00000003137	CYP26B1	-0.264
ENSG00000149182	ARFGAP2	-0.261

Table B.1 (continued)

ENSG00000204843	DCTN1	-0.261
ENSG00000168952	STXBP6	-0.26
ENSG00000198898	CAPZA2	-0.26
ENSG00000149575	SCN2B	-0.249
ENSG00000139597	N4BP2L1	-0.246
ENSG00000160948	VPS28	-0.243
ENSG00000102109	PCSK1N	-0.243
ENSG00000204764	RANBP17	-0.243
ENSG00000113851	CRBN	-0.24
ENSG00000120820	GLT8D2	-0.239
ENSG00000164252	AGGF1	-0.237
ENSG00000152669	CCNO	-0.23
ENSG00000075239	ACAT1	-0.222
ENSG00000172172	MRPL13	-0.22
ENSG00000136003	ISCU	-0.22
ENSG00000149100	EIF3M	-0.218
ENSG00000159363	ATP13A2	-0.211
ENSG00000167005	NUDT21	-0.211
ENSG00000100285	NEFH	-0.21
ENSG00000029534	ANK1	-0.21
ENSG00000211460	TSN	-0.208
ENSG00000161281	COX7A1	-0.197
ENSG00000198131	ZNF544	-0.196
ENSG00000127588	GNG13	-0.192
ENSG00000101843	PSMD10	-0.188
ENSG00000151876	FBXO4	-0.18
ENSG00000089050	RBBP9	-0.157
ENSG00000117419	ERI3	-0.127
ENSG00000087085	ACHE	-0.12

APPENDIX C

LIST OF GO BIOLOGICAL PROCESS CATEGORIES ENRICHED IN UPREGULATED GENES

Table C.1: List of significant GO BP Categories having OR>1 for upregulated vs. downregulated common genes across datasets. OR is the odds ratio calculated by FET. *p-adj* is the *BY* adjusted p value.

GO ID	GO Term	OR	*p-adj*
GO:0000122	negative regulation of transcription fro..	4.71	0
GO:0001501	skeletal system development..	3.74	0.003
GO:0001503	ossification..	4.28	0.004
GO:0001525	angiogenesis..	3.61	0.011
GO:0001568	blood vessel development..	4.66	0
GO:0001655	urogenital system development..	6.75	0
GO:0001656	metanephros development..	12.98	0.044
GO:0001657	ureteric bud development..	7.91	0.036
GO:0001701	in utero embryonic development..	4.53	0.002
GO:0001763	morphogenesis of a branching structure..	6.87	0.018
GO:0001775	cell activation..	2.32	0.026
GO:0001816	cytokine production..	4.04	0
GO:0001817	regulation of cytokine production..	4.22	0
GO:0001819	positive regulation of cytokine producti..	3.41	0.05
GO:0001822	kidney development..	8.88	0
GO:0001823	mesonephros development..	8.39	0.024
GO:0001890	placenta development..	13.93	0.027
GO:0001892	embryonic placenta development..	Inf	0.026
GO:0001932	regulation of protein phosphorylation..	1.88	0.026

Table C.1 (continued)

GO:0001934	positive regulation of protein phosphory..	2.14	0.018
GO:0001942	hair follicle development..	Inf	0.016
GO:0001944	vasculature development..	4.21	0
GO:0002009	morphogenesis of an epithelium..	3.51	0.004
GO:0002064	epithelial cell development..	11.29	0.001
GO:0002221	pattern recognition receptor signaling p..	9.35	0.01
GO:0002224	toll-like receptor signaling pathway..	7.43	0.057
GO:0002237	response to molecule of bacterial origin..	4.33	0.009
GO:0002250	adaptive immune response..	4.69	0.027
GO:0002252	immune effector process..	2.91	0.002
GO:0002253	activation of immune response..	3	0.014
GO:0002376	immune system process..	2.66	0
GO:0002520	immune system development..	4.03	0
GO:0002521	leukocyte differentiation..	2.75	0.088
GO:0002682	regulation of immune system process..	3.07	0
GO:0002683	negative regulation of immune system pro..	8.88	0
GO:0002684	positive regulation of immune system pro..	3	0
GO:0002697	regulation of immune effector process..	3.31	0.029
GO:0002764	immune response-regulating signaling pat..	2.64	0.024
GO:0003002	regionalization..	6.12	0.008
GO:0003006	developmental process involved in reprod..	2.85	0.026
GO:0003007	heart morphogenesis..	4.39	0.019
GO:0003158	endothelium development..	15.84	0.011
GO:0006366	transcription from RNA polymerase II pro..	3.01	0
GO:0051726	regulation of cell cycle..	2.23	0.018
GO:0006952	defense response..	2.2	0.001
GO:0006606	protein import into nucleus..	4.39	0.019
GO:0007010	cytoskeleton organization..	2.08	0.032
GO:0006915	apoptotic process..	2.53	0
GO:0007155	cell adhesion..	2.37	0.001
GO:0007165	signal transduction..	1.49	0.041
GO:0035556	intracellular signal transduction..	1.65	0.028
GO:0007178	transmembrane receptor protein serine/th..	4.18	0.013
GO:0016570	histone modification..	4.3	0.062
GO:0006325	chromatin organization..	3.1	0.036
GO:0006351	transcription, DNA-templated..	2.69	0
GO:0006355	regulation of transcription, DNA-templat..	2.97	0
GO:0006357	regulation of transcription from RNA pol..	4	0

Table C.1 (continued)

GO:0006468	protein phosphorylation..	1.79	0.024
GO:0006950	response to stress..	1.88	0
GO:0006954	inflammatory response..	2.68	0.013
GO:0006955	immune response..	2.66	0
GO:0006959	humoral immune response..	13.93	0.027
GO:0007049	cell cycle..	1.86	0.02
GO:0040007	growth..	2.14	0.011
GO:0007166	cell surface receptor signaling pathway..	1.86	0.001
GO:0007229	integrin-mediated signaling pathway..	15.84	0.011
GO:0007249	I-kappaB kinase/NF-kappaB signaling..	4.04	0.018
GO:0007275	multicellular organismal development..	1.74	0.001
GO:0009653	anatomical structure morphogenesis..	1.7	0.011
GO:0009790	embryo development..	2.91	0
GO:0007389	pattern specification process..	4.33	0.009
GO:0009887	organ morphogenesis..	4.36	0
GO:0009888	tissue development..	3.02	0
GO:0007423	sensory organ development..	2.59	0.056
GO:0007498	mesoderm development..	12.98	0.044
GO:0007507	heart development..	3.07	0.006
GO:0043066	negative regulation of apoptotic process..	2.68	0.002
GO:0008219	cell death..	2.58	0
GO:0012501	programmed cell death..	2.51	0
GO:0008283	cell proliferation..	2.97	0
GO:0008284	positive regulation of cell proliferatio..	2.82	0.003
GO:0008285	negative regulation of cell proliferatio..	2.61	0.005
GO:0008544	epidermis development..	4.89	0.019
GO:0008630	intrinsic apoptotic signaling pathway in..	13.93	0.027
GO:0009058	biosynthetic process..	1.52	0.027
GO:0009059	macromolecule biosynthetic process..	2.1	0
GO:0009605	response to external stimulus..	1.6	0.064
GO:0009607	response to biotic stimulus..	3.03	0.001
GO:0009611	response to wounding..	2.23	0.006
GO:0051707	response to other organism..	2.92	0.001
GO:0009617	response to bacterium..	3.72	0.007
GO:0009792	embryo development ending in birth or eg..	3.2	0.003
GO:0009889	regulation of biosynthetic process..	2.48	0
GO:0009890	negative regulation of biosynthetic proc..	2.96	0
GO:0009891	positive regulation of biosynthetic proc..	3.18	0

Table C.1 (continued)

GO:0009892	negative regulation of metabolic process..	1.89	0.002
GO:0009893	positive regulation of metabolic process..	2.35	0
GO:0009952	anterior/posterior pattern specification..	9.83	0.006
GO:0009966	regulation of signal transduction..	1.87	0.002
GO:0009967	positive regulation of signal transducti..	2.5	0
GO:0010033	response to organic substance..	1.57	0.044
GO:0010467	gene expression..	1.87	0
GO:0010468	regulation of gene expression..	2.73	0
GO:0010556	regulation of macromolecule biosynthetic..	2.75	0
GO:0010557	positive regulation of macromolecule bio..	3.49	0
GO:0010558	negative regulation of macromolecule bio..	3.34	0
GO:0010562	positive regulation of phosphorus metabo..	1.87	0.06
GO:0010604	positive regulation of macromolecule met..	2.41	0
GO:0010605	negative regulation of macromolecule met..	2.09	0
GO:0010628	positive regulation of gene expression..	3.18	0
GO:0010629	negative regulation of gene expression..	3.89	0
GO:0010647	positive regulation of cell communicatio..	1.74	0.05
GO:0048646	anatomical structure formation involved ..	2.17	0.01
GO:0010941	regulation of cell death..	2.54	0
GO:0010942	positive regulation of cell death..	2.44	0.035
GO:0014031	mesenchymal cell development..	12.98	0.044
GO:0016070	RNA metabolic process..	2.03	0
GO:0045935	positive regulation of nucleobase-contai..	3.24	0
GO:0016265	death..	2.58	0
GO:0042981	regulation of apoptotic process..	2.6	0
GO:0016477	cell migration..	3.15	0
GO:0045893	positive regulation of transcription, DN..	3.71	0
GO:0045892	negative regulation of transcription, DN..	4.64	0
GO:0016568	chromatin modification..	4.33	0.009
GO:0016569	covalent chromatin modification..	4.3	0.062
GO:0017038	protein import..	3.02	0.094
GO:0018108	peptidyl-tyrosine phosphorylation..	3.41	0.05
GO:0018130	heterocycle biosynthetic process..	2.17	0
GO:0018212	peptidyl-tyrosine modification..	3.41	0.05
GO:0019219	regulation of nucleobase-containing comp..	2.51	0
GO:0019222	regulation of metabolic process..	2.04	0
GO:0019438	aromatic compound biosynthetic process..	2.24	0
GO:0019827	stem cell population maintenance..	5.63	0.019

Table C.1 (continued)

GO:0022404	molting cycle process..	Inf	0.016
GO:0022405	hair cycle process..	Inf	0.016
GO:0022414	reproductive process..	2.42	0.005
GO:0022603	regulation of anatomical structure morph..	2.16	0.04
GO:0022610	biological adhesion..	2.43	0.001
GO:0070887	cellular response to chemical stimulus..	1.62	0.029
GO:0023056	positive regulation of signaling..	1.7	0.076
GO:0030097	hemopoiesis..	3.58	0
GO:0030099	myeloid cell differentiation..	3.48	0.079
GO:0030154	cell differentiation..	1.84	0
GO:0030155	regulation of cell adhesion..	3	0.021
GO:0030198	extracellular matrix organization..	6.68	0
GO:0030278	regulation of ossification..	6.54	0.018
GO:0030323	respiratory tube development..	4.49	0.042
GO:0030324	lung development..	4.11	0.095
GO:0030326	embryonic limb morphogenesis..	7.43	0.057
GO:0030334	regulation of cell migration..	3.08	0.015
GO:0030335	positive regulation of cell migration..	3.28	0.072
GO:0030509	BMP signaling pathway..	13.93	0.027
GO:0030855	epithelial cell differentiation..	7.1	0
GO:0030856	regulation of epithelial cell differenti..	16.8	0.007
GO:0031323	regulation of cellular metabolic process..	2	0
GO:0031324	negative regulation of cellular metaboli..	1.97	0.002
GO:0031325	positive regulation of cellular metaboli..	2.49	0
GO:0031326	regulation of cellular biosynthetic proc..	2.4	0
GO:0031327	negative regulation of cellular biosynth..	2.99	0
GO:0031328	positive regulation of cellular biosynth..	3.06	0
GO:0031401	positive regulation of protein modificat..	1.98	0.026
GO:0031589	cell-substrate adhesion..	6.08	0.001
GO:0032101	regulation of response to external stimu..	2.13	0.056
GO:0032270	positive regulation of cellular protein ..	1.93	0.014
GO:0032496	response to lipopolysaccharide..	4.18	0.013
GO:0032501	multicellular organismal process..	1.67	0.002
GO:0032502	developmental process..	1.8	0
GO:0032774	RNA biosynthetic process..	2.61	0
GO:0032835	glomerulus development..	Inf	0.043
GO:0034329	cell junction assembly..	4.23	0.027
GO:0034330	cell junction organization..	4.39	0.019

Table C.1 (continued)

GO:0034504	protein localization to nucleus..	5.04	0.001
GO:0034645	cellular macromolecule biosynthetic proc..	2.09	0
GO:0034654	nucleobase-containing compound biosynthe..	2.29	0
GO:0035107	appendage morphogenesis..	9.35	0.01
GO:0035108	limb morphogenesis..	9.35	0.01
GO:0035113	embryonic appendage morphogenesis..	7.43	0.057
GO:0035239	tube morphogenesis..	3.41	0.021
GO:0035295	tube development..	3.49	0
GO:0040011	locomotion..	1.71	0.067
GO:0040012	regulation of locomotion..	2.54	0.036
GO:0042060	wound healing..	2.13	0.062
GO:0042127	regulation of cell proliferation..	2.95	0
GO:0042303	molting cycle..	Inf	0.006
GO:0042325	regulation of phosphorylation..	1.86	0.029
GO:0042327	positive regulation of phosphorylation..	2.11	0.019
GO:0042476	odontogenesis..	Inf	0.026
GO:0042633	hair cycle..	Inf	0.006
GO:0042692	muscle cell differentiation..	3.31	0.029
GO:0042733	embryonic digit morphogenesis..	Inf	0.026
GO:0043009	chordate embryonic development..	3.35	0.001
GO:0043062	extracellular structure organization..	6.68	0
GO:0043065	positive regulation of apoptotic process..	2.54	0.036
GO:0043067	regulation of programmed cell death..	2.63	0
GO:0043068	positive regulation of programmed cell d..	2.6	0.027
GO:0043069	negative regulation of programmed cell d..	2.72	0.001
GO:0043122	regulation of I-kappaB kinase/NF-kappaB ..	4.56	0.013
GO:0043123	positive regulation of I-kappaB kinase/N..	5.63	0.019
GO:0043170	macromolecule metabolic process..	1.55	0.016
GO:0043207	response to external biotic stimulus..	2.92	0.001
GO:0043588	skin development..	5.38	0.027
GO:0043900	regulation of multi-organism process..	3.24	0.045
GO:0044093	positive regulation of molecular functio..	1.76	0.049
GO:0044249	cellular biosynthetic process..	1.56	0.016
GO:0044260	cellular macromolecule metabolic process..	1.46	0.054
GO:0044271	cellular nitrogen compound biosynthetic ..	1.9	0
GO:0044702	single organism reproductive process..	2.42	0.011
GO:0044707	single-multicellular organism process..	1.69	0.001
GO:0044744	protein targeting to nucleus..	4.39	0.019

Table C.1 (continued)

GO:0044767	single-organism developmental process..	1.77	0
GO:0045087	innate immune response..	2.29	0.021
GO:0045165	cell fate commitment..	6.22	0.027
GO:0045446	endothelial cell differentiation..	12.98	0.044
GO:0045595	regulation of cell differentiation..	2.59	0
GO:0045596	negative regulation of cell differentiat..	3.52	0.001
GO:0045597	positive regulation of cell differentiat..	2.3	0.009
GO:0045765	regulation of angiogenesis..	6.54	0.018
GO:0045785	positive regulation of cell adhesion..	4.23	0.027
GO:0045934	negative regulation of nucleobase-contai..	3.07	0
GO:0045937	positive regulation of phosphate metabol..	1.87	0.06
GO:0045944	positive regulation of transcription fro..	4.7	0
GO:0048513	organ development..	2.82	0
GO:0048514	blood vessel morphogenesis..	3.86	0.001
GO:0048518	positive regulation of biological proces..	2.1	0
GO:0048519	negative regulation of biological proces..	1.95	0
GO:0048522	positive regulation of cellular process..	1.97	0
GO:0048523	negative regulation of cellular process..	1.94	0
GO:0048534	hematopoietic or lymphoid organ developm..	3.64	0
GO:0048565	digestive tract development..	6.87	0.018
GO:0048568	embryonic organ development..	4.66	0.001
GO:0048583	regulation of response to stimulus..	1.96	0
GO:0048584	positive regulation of response to stimu..	2.48	0
GO:0048598	embryonic morphogenesis..	2.7	0.018
GO:0048608	reproductive structure development..	3.8	0.008
GO:0048705	skeletal system morphogenesis..	6.95	0.091
GO:0048729	tissue morphogenesis..	3.82	0
GO:0048731	system development..	1.66	0.004
GO:0048732	gland development..	5.22	0.002
GO:0048736	appendage development..	6.87	0.018
GO:0048754	branching morphogenesis of an epithelial..	6.22	0.027
GO:0048762	mesenchymal cell differentiation..	12.98	0.044
GO:0048856	anatomical structure development..	1.68	0.002
GO:0048863	stem cell differentiation..	20.67	0.001
GO:0048864	stem cell development..	12.98	0.044
GO:0048869	cellular developmental process..	1.76	0.001
GO:0048870	cell motility..	2.92	0
GO:0048872	homeostasis of number of cells..	8.39	0.024

Table C.1 (continued)

GO:0050673	epithelial cell proliferation..	3.62	0.054
GO:0050776	regulation of immune response..	3.14	0
GO:0050778	positive regulation of immune response..	2.78	0.011
GO:0050789	regulation of biological process..	1.71	0.003
GO:0050793	regulation of developmental process..	2.64	0
GO:0050794	regulation of cellular process..	1.62	0.009
GO:0050867	positive regulation of cell activation..	4.11	0.095
GO:0050896	response to stimulus..	1.56	0.015
GO:0050900	leukocyte migration..	9.83	0.006
GO:0051090	regulation of sequence-specific DNA bind..	3.69	0.012
GO:0051093	negative regulation of developmental pro..	3.52	0
GO:0051094	positive regulation of developmental pro..	2.25	0.003
GO:0051147	regulation of muscle cell differentiatio..	9.83	0.006
GO:0051148	negative regulation of muscle cell diffe..	Inf	0.076
GO:0051169	nuclear transport..	2.69	0.076
GO:0051170	nuclear import..	4.39	0.019
GO:0051171	regulation of nitrogen compound metaboli..	2.19	0
GO:0051172	negative regulation of nitrogen compound..	2.79	0
GO:0051173	positive regulation of nitrogen compound..	3.06	0
GO:0051216	cartilage development..	5.9	0.042
GO:0051239	regulation of multicellular organismal p..	2.03	0
GO:0051240	positive regulation of multicellular org..	1.88	0.021
GO:0051241	negative regulation of multicellular org..	2.49	0.002
GO:0051246	regulation of protein metabolic process..	1.72	0.012
GO:0051247	positive regulation of protein metabolic..	1.95	0.009
GO:0051252	regulation of RNA metabolic process..	2.88	0
GO:0051253	negative regulation of RNA metabolic pro..	4.08	0
GO:0051254	positive regulation of RNA metabolic pro..	3.79	0
GO:0051272	positive regulation of cellular componen..	3.53	0.035
GO:0051674	localization of cell..	2.92	0
GO:0051704	multi-organism process..	1.94	0.002
GO:0051716	cellular response to stimulus..	1.48	0.045
GO:0070848	response to growth factor..	2.15	0.02
GO:0055123	digestive system development..	5.14	0.043
GO:0060173	limb development..	6.87	0.018
GO:0060255	regulation of macromolecule metabolic pr..	2.2	0
GO:0060429	epithelium development..	3.59	0
GO:0060485	mesenchyme development..	6.22	0.027

Table C.1 (continued)

GO:0060541	respiratory system development..	4.3	0.062
GO:0060548	negative regulation of cell death..	2.6	0.002
GO:0060562	epithelial tube morphogenesis..	3.11	0.042
GO:0060993	kidney morphogenesis..	13.93	0.027
GO:0061005	cell differentiation involved in kidney ..	Inf	0.076
GO:0061061	muscle structure development..	2.35	0.076
GO:0061138	morphogenesis of a branching epithelium..	6.87	0.018
GO:0061326	renal tubule development..	6.95	0.091
GO:0061458	reproductive system development..	3.8	0.008
GO:0065007	biological regulation..	1.74	0.004
GO:0071216	cellular response to biotic stimulus..	9.83	0.006
GO:0071219	cellular response to molecule of bacteri..	8.87	0.016
GO:0071222	cellular response to lipopolysaccharide..	8.87	0.016
GO:0071310	cellular response to organic substance..	1.71	0.018
GO:0071363	cellular response to growth factor stimu..	2.02	0.042
GO:0071559	response to transforming growth factor b..	4.49	0.042
GO:0071772	response to BMP..	13.93	0.027
GO:0071773	cellular response to BMP stimulus..	13.93	0.027
GO:0071902	positive regulation of protein serine/th..	3.65	0.025
GO:0072001	renal system development..	7.91	0
GO:0072006	nephron development..	9.83	0.006
GO:0072009	nephron epithelium development..	8.39	0.024
GO:0072073	kidney epithelium development..	5.38	0.027
GO:0072163	mesonephric epithelium development..	7.91	0.036
GO:0072164	mesonephric tubule development..	7.91	0.036
GO:0072358	cardiovascular system development..	3.26	0
GO:0072359	circulatory system development..	3.26	0
GO:0080090	regulation of primary metabolic process..	2.05	0
GO:0080134	regulation of response to stress..	2.22	0.001
GO:0090100	positive regulation of transmembrane rec..	12.98	0.044
GO:0090304	nucleic acid metabolic process..	1.91	0
GO:0097190	apoptotic signaling pathway..	2.35	0.035
GO:0097659	nucleic acid-templated transcription..	2.67	0
GO:0098542	defense response to other organism..	3.09	0.024
GO:0098602	single organism cell adhesion..	2.32	0.054
GO:0098727	maintenance of cell number..	4.49	0.042
GO:0098773	skin epidermis development..	Inf	0.016
GO:1901342	regulation of vasculature development..	6.87	0.018

Table C.1 (continued)

GO:1901362	organic cyclic compound biosynthetic pro..	2.15	0
GO:1901576	organic substance biosynthetic process..	1.54	0.023
GO:1902106	negative regulation of leukocyte differe..	Inf	0.043
GO:1902531	regulation of intracellular signal trans..	1.7	0.076
GO:1902533	positive regulation of intracellular sig..	2.32	0.006
GO:1902593	single-organism nuclear import..	4.39	0.019
GO:1902679	negative regulation of RNA biosynthetic ..	4.21	0
GO:1902680	positive regulation of RNA biosynthetic ..	3.6	0
GO:1903506	regulation of nucleic acid-templated tra..	2.93	0
GO:1903507	negative regulation of nucleic acid-temp..	4.15	0
GO:1903508	positive regulation of nucleic acid-temp..	3.71	0
GO:1903706	regulation of hemopoiesis..	4.89	0.006
GO:1903707	negative regulation of hemopoiesis..	Inf	0.002
GO:2000026	regulation of multicellular organismal d..	2.23	0
GO:2000112	regulation of cellular macromolecule bio..	2.76	0
GO:2000113	negative regulation of cellular macromol..	3.76	0
GO:2000145	regulation of cell motility..	2.74	0.02
GO:2000147	positive regulation of cell motility..	3.28	0.072
GO:2000736	regulation of stem cell differentiation..	13.93	0.027
GO:2001141	regulation of RNA biosynthetic process..	2.93	0

APPENDIX D

LIST OF GO BIOLOGICAL PROCESS CATEGORIES ENRICHED IN DOWNREGULATED GENES

Table D.1: List of significant GO BP Categories having OR<1 for upregulated vs. downregulated common genes across datasets. OR is the odds ratio calculated by FET. *p-adj* is the *BY* adjusted p value.

GO:0019941	modification-dependent protein catabolic..	0.31	0.033
GO:0042787	protein ubiquitination involved in ubiqu..	0.13	0.058
GO:0001505	regulation of neurotransmitter levels..	0.17	0.012
GO:0002790	peptide secretion..	0.25	0.059
GO:0002791	regulation of peptide secretion..	0.2	0.066
GO:0007267	cell-cell signaling..	0.32	0
GO:0023061	signal release..	0.24	0.001
GO:0006091	generation of precursor metabolites and ..	0.28	0.001
GO:0022904	respiratory electron transport chain..	0.04	0.001
GO:0006119	oxidative phosphorylation..	0.07	0.027
GO:0022900	electron transport chain..	0.04	0.001
GO:0006163	purine nucleotide metabolic process..	0.32	0.007
GO:0042278	purine nucleoside metabolic process..	0.31	0.015
GO:0006414	translational elongation..	0.08	0.079
GO:0051603	proteolysis involved in cellular protein..	0.37	0.061
GO:0006511	ubiquitin-dependent protein catabolic pr..	0.31	0.033
GO:0006732	coenzyme metabolic process..	0.22	0.007
GO:0006733	oxidoreduction coenzyme metabolic proces..	0.15	0.026
GO:0006753	nucleoside phosphate metabolic process..	0.3	0.001
GO:0006810	transport..	0.69	0.079

Table D.1 (continued)

GO:0006811	ion transport..	0.48	0.003
GO:0006812	cation transport..	0.51	0.049
GO:0006818	hydrogen transport..	0.08	0.002
GO:0006836	neurotransmitter transport..	0.22	0.018
GO:0007005	mitochondrion organization..	0.4	0.027
GO:0007215	glutamate receptor signaling pathway..	0	0.009
GO:0007268	synaptic transmission..	0.16	0
GO:0007269	neurotransmitter secretion..	0.19	0.043
GO:0007270	neuron-neuron synaptic transmission..	0.07	0
GO:0050877	neurological system process..	0.36	0.001
GO:0007610	behavior..	0.43	0.013
GO:0007611	learning or memory..	0.21	0.005
GO:0007612	learning..	0.16	0.041
GO:0009108	coenzyme biosynthetic process..	0.12	0.035
GO:0009116	nucleoside metabolic process..	0.31	0.015
GO:0009117	nucleotide metabolic process..	0.31	0.002
GO:0009119	ribonucleoside metabolic process..	0.31	0.015
GO:0009123	nucleoside monophosphate metabolic proce..	0.22	0.004
GO:0009126	purine nucleoside monophosphate metaboli..	0.22	0.004
GO:0009141	nucleoside triphosphate metabolic proces..	0.25	0.01
GO:0009144	purine nucleoside triphosphate metabolic..	0.27	0.02
GO:0009150	purine ribonucleotide metabolic process..	0.32	0.007
GO:0009161	ribonucleoside monophosphate metabolic p..	0.22	0.004
GO:0009167	purine ribonucleoside monophosphate meta..	0.22	0.004
GO:0009199	ribonucleoside triphosphate metabolic pr..	0.27	0.02
GO:0009205	purine ribonucleoside triphosphate metab..	0.27	0.02
GO:0009259	ribonucleotide metabolic process..	0.32	0.007
GO:0009914	hormone transport..	0.28	0.047
GO:0010498	proteasomal protein catabolic process..	0.33	0.079
GO:0015672	monovalent inorganic cation transport..	0.33	0.002
GO:0015980	energy derivation by oxidation of organi..	0.3	0.006
GO:0015992	proton transport..	0.08	0.002
GO:0016358	dendrite development..	0.2	0.066
GO:0019637	organophosphate metabolic process..	0.53	0.072
GO:0019693	ribose phosphate metabolic process..	0.32	0.007
GO:0030072	peptide hormone secretion..	0.21	0.027
GO:0032543	mitochondrial translation..	0	0.009
GO:0034220	ion transmembrane transport..	0.38	0.001

Table D.1 (continued)

GO:0035249	synaptic transmission, glutamatergic..	0.08	0.079
GO:0042391	regulation of membrane potential..	0.12	0
GO:0042773	ATP synthesis coupled electron transport..	0.08	0.079
GO:0042775	mitochondrial ATP synthesis coupled elec..	0.08	0.079
GO:0043161	proteasome-mediated ubiquitin-dependent ..	0.32	0.091
GO:0043632	modification-dependent macromolecule cat..	0.31	0.033
GO:0043648	dicarboxylic acid metabolic process..	0	0.009
GO:0044281	small molecule metabolic process..	0.59	0.023
GO:0044708	single-organism behavior..	0.25	0
GO:0044765	single-organism transport..	0.64	0.025
GO:0045333	cellular respiration..	0.06	0
GO:0046034	ATP metabolic process..	0.21	0.005
GO:0046128	purine ribonucleoside metabolic process..	0.31	0.015
GO:0046879	hormone secretion..	0.29	0.07
GO:0046883	regulation of hormone secretion..	0.17	0.012
GO:0048167	regulation of synaptic plasticity..	0.05	0.002
GO:0050803	regulation of synapse structure or activ..	0.11	0.002
GO:0050804	modulation of synaptic transmission..	0.11	0
GO:0050806	positive regulation of synaptic transmis..	0.12	0.035
GO:0050890	cognition..	0.23	0.007
GO:0051186	cofactor metabolic process..	0.2	0.003
GO:0051188	cofactor biosynthetic process..	0.1	0.014
GO:0051648	vesicle localization..	0.29	0.07
GO:0051932	synaptic transmission, GABAergic..	0.08	0.079
GO:0055085	transmembrane transport..	0.46	0.003
GO:0055086	nucleobase-containing small molecule met..	0.35	0.004
GO:0070125	mitochondrial translational elongation..	0	0.027
GO:0072521	purine-containing compound metabolic pro..	0.38	0.031
GO:0090087	regulation of peptide transport..	0.27	0.091
GO:0090276	regulation of peptide hormone secretion..	0.2	0.066
GO:0098655	cation transmembrane transport..	0.28	0
GO:0098660	inorganic ion transmembrane transport..	0.29	0
GO:0098662	inorganic cation transmembrane transport..	0.27	0
GO:1901564	organonitrogen compound metabolic proces..	0.54	0.01
GO:1901657	glycosyl compound metabolic process..	0.31	0.015
GO:1902578	single-organism localization..	0.67	0.054
GO:1902600	hydrogen ion transmembrane transport..	0.09	0.005
GO:1903052	positive regulation of proteolysis invol..	0.13	0.058