

Washington University in St. Louis

## Washington University Open Scholarship

---

All Computer Science and Engineering  
Research

Computer Science and Engineering

---

Report Number: WUCSE-2007-36

2007

### Single cell expression profiling reveals major disruption of DNA repair capacity in incipient Alzheimer's Disease

Monika Ray and Weixiong Zhang

Understanding the pathogenesis in the early stages of late-onset Alzheimer's disease (LOAD) can help in gaining important mechanistic insights into this devastating neurodegenerative disorder. Alzheimer's disease (AD) is characterised by extensive cell death with disease progression. In this paper laser capture microdissection (LCM) based gene expression profiling, which is able to profile gene expression in a single cell type, is employed to analyse the gene expression regulation of incipient AD. Our analysis shows that LCM based gene expression profiling of neurons has a critical advantage over the conventional gene expression profiling method which uses samples of mixed cell types... [Read complete abstract on page 2.](#)

Follow this and additional works at: [https://openscholarship.wustl.edu/cse\\_research](https://openscholarship.wustl.edu/cse_research)



Part of the [Computer Engineering Commons](#), and the [Computer Sciences Commons](#)

---

#### Recommended Citation

Ray, Monika and Zhang, Weixiong, "Single cell expression profiling reveals major disruption of DNA repair capacity in incipient Alzheimer's Disease" Report Number: WUCSE-2007-36 (2007). *All Computer Science and Engineering Research*.

[https://openscholarship.wustl.edu/cse\\_research/137](https://openscholarship.wustl.edu/cse_research/137)

Department of Computer Science & Engineering - Washington University in St. Louis  
Campus Box 1045 - St. Louis, MO - 63130 - ph: (314) 935-6160.

## Single cell expression profiling reveals major disruption of DNA repair capacity in incipient Alzheimer's Disease

Monika Ray and Weixiong Zhang

### Complete Abstract:

Understanding the pathogenesis in the early stages of late-onset Alzheimer's disease (LOAD) can help in gaining important mechanistic insights into this devastating neurodegenerative disorder. Alzheimer's disease (AD) is characterised by extensive cell death with disease progression. In this paper laser capture microdissection (LCM) based gene expression profiling, which is able to profile gene expression in a single cell type, is employed to analyse the gene expression regulation of incipient AD. Our analysis shows that LCM based gene expression profiling of neurons has a critical advantage over the conventional gene expression profiling method which uses samples of mixed cell types and does not account for cell loss. A natural question that arises in this situation is how can we say that a gene is "down-regulated" when it just may not be present due to the absence of cells? Our investigation on expression regulation also resulted in 1633 significant genes, from which we identified 15 DNA repair genes that had elevated or lowered levels of mRNA expression. Many key players involved in the defense against DNA damage were present in this list, and many genes were involved in the base-excision repair pathway. Our results indicate that the base excision repair pathway, a predominant pathway in neurons, fails to handle the DNA damage in an efficient manner and is a compelling explanation for many of the pathological features of LOAD.

2007-36

## Single cell expression profiling reveals major disruption of DNA repair capacity in incipient Alzheimer's Disease

Authors: Monika Ray and Weixiong Zhang

Corresponding Author: [mray@cse.wustl.edu](mailto:mray@cse.wustl.edu)

**Abstract:** Understanding the pathogenesis in the early stages of late-onset Alzheimer's disease (LOAD) can help in gaining important mechanistic insights into this devastating neurodegenerative disorder.

Alzheimer's disease (AD) is characterised by extensive cell death with disease progression. In this paper laser capture microdissection (LCM) based gene expression profiling, which is able to profile gene expression in a single cell type, is employed to analyse the gene expression regulation of incipient AD. Our analysis shows that LCM based gene expression profiling of neurons has a critical advantage over the conventional gene expression profiling method which uses samples of mixed cell types and does not account for cell loss. A natural question that arises in this situation is how can we say that a gene is "down-regulated" when it just may not be present due to the absence of cells? Our investigation on expression regulation also resulted in 1633 significant genes, from which we identified 15 DNA repair genes that had elevated or lowered levels of mRNA expression. Many key players involved in the defense against DNA damage were present in this list, and many genes were involved in the base-excision repair pathway. Our results indicate that the base excision repair pathway, a predominant pathway in neurons, fails to handle the DNA damage in an efficient manner and is a compelling explanation for many of the pathological features of LOAD.

# Single cell expression profiling reveals major disruption of DNA repair capacity in incipient Alzheimer's Disease

Monika Ray<sup>1</sup>, Weixiong Zhang<sup>1,2\*</sup>

<sup>1</sup>Washington University School of Engineering,

Dept. of Computer Science and Engineering, St. Louis, MO 63130

<sup>2</sup>Washington University School of Medicine,

Dept. of Genetics, St. Louis, MO 63110

\* Corresponding author

## Abstract

Understanding the pathogenesis in the early stages of late-onset Alzheimer's disease (LOAD) can help in gaining important mechanistic insights into this devastating neurodegenerative disorder. Alzheimer's disease (AD) is characterised by extensive cell death with disease progression. In this paper laser capture microdissection (LCM) based gene expression profiling, which is able to profile gene expression in a single cell type, is employed to analyse the gene expression regulation of incipient AD. Our analysis shows that LCM based gene expression profiling of neurons has a critical advantage over the conventional gene expression profiling method which uses samples of mixed cell types and does not account for cell loss. A natural question that arises in this situation is how can we say that a gene is "down-regulated" when it just may not be present due to the absence of cells? Our investigation on expression regulation also resulted in 1633 significant genes, from which we identified 15 DNA repair genes that had elevated or lowered levels of mRNA expression. Many key players involved in the defense against DNA damage were present in this list, and many genes were involved in the base-excision repair pathway. Our results indicate that the base excision repair pathway, a predominant pathway in neurons, fails to handle the DNA damage in an efficient manner and is a compelling explanation for many of the pathological features of LOAD.

## Introduction

Late-onset Alzheimer's disease (LOAD) is a complex progressive neurodegenerative disorder of the brain and is the commonest form of dementia. Advancing age is the major contributing factor for increased susceptibility to Alzheimer's Disease (AD) and the old are the fastest-growing segment of the United States population and have the highest prevalence of dementia [1]. Early diagnosis of AD can help in effective treatment strategies. Therefore, studying the pathogenesis of incipient AD is very crucial.

Pathologically AD is characterised by the presence of neurofibrillary tangles (NFT) in the neurons of the cerebral cortex and hippocampus. Six stages of AD disease propagation have been distinguished: transentorhinal stages I-II (clinically silent cases), limbic stages III-IV (incipient AD) and neocortical stages V-VI (fully developed AD) [2]. The neurofibrillary pathology begins at the transentorhinal and entorhinal cortices and then proceeds to the hippocampus [3]. AD is also characterised by marked neuronal cell loss. In particular cell loss has been correlated with the amount of NFT but not with age or  $\beta$ /A4-amyloid deposition [4]. Furthermore, it has been shown that neuronal cell loss is greater than volume loss (51% and 25%, respectively) between stages I and V in the hippocampus [2].

Tissues are structures with many different cell types. In disease pathologies, the diseased cells of interest are surrounded by these heterogeneous tissue elements. Cells that are primarily affected in a disease and depict molecular changes due to disease progression may constitute a very small fraction of the volume of the tissue biopsy sample. In the last few years, there have been many gene expression studies related to AD. The majority of them have used samples obtained via regular dissection from affected brain regions. Considering that there is neuronal cell loss (even greater than volume loss) in neurodegenerative diseases and the heterogeneity of samples due to regular dissection, interpretation of gene expression results may be difficult and somewhat ambiguous. Variability in gene expression may not be due to AD but rather resulting from cellular heterogeneity of brain tissues. Furthermore, since there is cell loss, how can it be concluded that the decrease or increase in mRNA expression levels is actually a "down-regulation" or "up-regulation" when it can also be due to the absence or presence of cells? Laser Capture Microdissection (LCM), which was introduced by the NIH in 1996 [5], is a method for procuring pure cells from specific microscopic regions of tissue sections. In this procedure, the kind and number of cells can be controlled so that one can compare the mRNA expression level of a gene in the same number of cells of the same type across different samples.

We have two main objectives in this study - the first is to compare Blalock et al.'s [6]

AD gene expression dataset obtained via regular tissue dissection with the LCM data obtained by Dunckley et al. [7] to distinguish the power of the dataset and interpretability of results. We perform the comparison by analysing the results of two differentially expressed gene selection methods - SAM [8] and Limma [9]- when applied to the two datasets. As microdissection is expensive and requires highly specialised technical ability, it is worthwhile to examine the difference, if any, between the two datasets obtained via these two tissue dissection methods. Secondly, although it has been suggested that DNA repair is compromised in AD, it has not been studied via microarray analysis in proper detail. In view of this aspect, the second aim of the present study is to mine Dunckley et al.'s data, and examine the DNA repair capacity in incipient AD. We feel that this dataset has not been thoroughly analysed by Dunckley et al. and a lot of potentially new knowledge regarding LOAD is present in the data. We want to investigate whether defects in the DNA repair pathways contribute to cell death that is typical in AD. In this article we report the activity of 15 DNA repair genes that were up or down-regulated in the AD individuals. Our analysis indicates a major disruption of the base-excision repair pathway as many of the significant genes identified were key players of this pathway.

To the best of our knowledge, this is the first study that compares the power of regular dissection and LCM-based gene expression profiling methods in a critical human disease. This is also the first study that has focused on the expression levels of DNA repair genes in LOAD affected individuals via microarray analysis. The investigation of genes perturbed in the DNA repair pathways via microarray analysis is important as it provides differentially expressed genes that are significant from a systems biology perspective and not in isolation from other genes. A gene may be significant when considered separately from the rest of the genes, however, when its effects are inspected taking all the other genes' activity into account, it may or may not be significantly differentially expressed.

## Results and Discussion

We used SAM [8] and Limma [9] to identify differentially expressed genes (See Methods).

### *Comparison of laser capture microdissection and regular dissection based gene expression datasets*

Table 1 shows the number of significant genes obtained from Blalock et al.'s and Dunckley et al.'s datasets using SAM and Limma. The false discovery rates (FDRs) shown in this table are the *lowest* FDRs achieved by each dataset using each of the 2 gene selection

methods. Although at a FDR of 1.1E-09 Limma gives 2 genes, at a FDR of 0.10%, Limma outputs 4372 significant genes. It is evident from these results that Dunckley et al.'s data provides a large number of differentially expressed transcripts when compared to Blalock et al.'s data, at reasonably low FDRs. These results clearly show the "richness" of Dunckley et al.'s dataset which was generated from microdissected samples.

In SAM, while the list of 1663 genes had up and down regulated genes, the list of 12 genes from Blalock et al.'s data only consisted of genes with decreased levels of expression. Other than the problem of providing few genes with low FDR in Blalock et al.'s data, there is also the issue with interpreting the 12 genes. The lowered levels of expression could be due to two reasons - (1)down-regulation of these genes in Braak stage V AD or (2)the absence of cells due to cell death resulting in no mRNA levels to measure. This dilemma is eliminated in Dunckley et al.'s data as cells of the same type, i.e. neurons, were obtained via LCM [7]. This means that one can compare the mRNA expression levels in a constant number of living cells across all samples. Hence, if a gene has lowered level of expression, it means it is down-regulated.

Another issue with Blalock et al.'s data is that it neither contains samples of AD with Braak stage III, which is incipient AD, nor is it taken from the entorhinal cortex, where AD pathology occurs first [10, 2]. Hence, analysis of Blalock et al.'s data may not be appropriate for studying the pathogenesis of incipient AD. As our aim is to examine expression changes in incipient AD, we further analyse the significant genes list obtained from Dunckley et al.'s data by SAM. We arbitrarily chose to analyse SAM's results over Limma's, although we believe that we would find similar results using Limma.

We believe that the aforementioned issues have been previously overlooked.

### ***Analysis of DNA repair genes in incipient AD - Results from Dunckley et al.'s LCM dataset***

EASE (<http://niaid.abcc.ncifcrf.gov/home.jsp>) was used to identify statistically significant biological processes present in the list of significant genes identified from Dunckley et al.'s dataset using SAM (See Methods). As Dunckley et al.'s dataset was generated from a 1000 neurons procured from each of the 33 samples (See Methods), the majority of biological processes are related to functions in neurons (see Table 2). One of the statistically significant biological process categories is DNA repair. Since the study of DNA repair in incipient AD is still in its nascent stage and has not been investigated in depth, we decided to inspect the set of genes involved in the defense against DNA damage.

Genomes are subject to damage by chemical and physical agents in the environment and by free radicals or alkylating agents endogenously generated during metabolism. Damage can result from hydrolysis, deamination, alkylation, and oxidation, all of which are capable of causing a modification in one or more bases in a DNA sequence. Mature neurons in the mammalian brain cannot divide and are highly metabolically active. Due to the high oxygen consumption rate by the brain, reactive oxygen species (ROS) can contribute to neuronal damage. Oxidative stress in neurons in human neurodegenerative diseases such as AD has been documented in previous reports [11, 12]. ROS attack of DNA can lead to DNA–DNA and DNA–protein cross linking, re-entry into cell-cycle by mature neurons, DNA strand breaks, production of oxidized base adducts, modification of DNA bases leading to problems in DNA replication and altered protein synthesis, and sister chromatid exchange and translocation in nuclear DNA [12]. The maintenance of genome integrity is essential and particularly important to neurons as they are among the longest living cells in the body. Accumulation of mutations and decline in DNA repair during disease progression can lead to cancerous conditions and/or cell death. In response to DNA damage, cells activate multiple signalling pathways, leading to the accumulation of proteins in complex multisubunit nuclear foci, that represent sites of DNA replication arrest or sites of DNA repair [13]. To deal with DNA damage, cells have evolved a repertoire of cell-cycle check-point and DNA repair processes. In order to repair DNA damage, four main DNA repair pathways are present - a simple reversal of the damage; nucleotide excision repair (NER), including mismatch and transcription-coupled repair; base excision repair (BER); and recombination repair including nonhomologous end joining [14, 15].

Table 3 shows the 15 DNA repair genes that were present in the list of 1663 significant genes from Dunckley et al.'s data. It can be seen that the majority of DNA repair genes have been significantly down-regulated.

Non-homologous end-joining (NHEJ) is the predominant pathway used to repair double-strand breaks in DNA and is evolutionarily conserved. Polynucleotide kinase 3'-phosphatase (PNKP) plays an important role in DNA double-strand breaks (DSBs) repair via NHEJ [16]. DSBs are the most lethal form of DNA damage [17]. A role of PNKP in single-strand break repair, along with DNA polymerase  $\beta$ , DNA ligase I, DNA ligase III and XRCC1, has also been identified [18, 19]. Deficiency of PNKP results in increased sensitivity to DNA damaging agents. PNKP is down-regulated in the AD samples by almost 2-fold.

NER is the most important repair process and removes the broadest spectrum of genomic damages, including UV-induced photoproducts, bulky mono-adducts, cross-links, and oxidative damage [20]. Several known genetic defects in NER lead to xeroderma pigmento-



sum, which is associated with a 1000-fold increase in skin cancer as well as a 20-fold increase in other internal tumors [21]. General transcription factor IIIH, polypeptide 1 (GTF2H1), which is decreased in AD samples by 1.5-fold, is a component of the core-TFIIH basal transcription factor involved in the NER of DNA [22].

Xeroderma pigmentosum, complementation group c (XPC) is up-regulated by 1.5-fold and aids in DNA damage recognition and in altering chromatin structure to allow access by damage-processing enzymes [22]. It recognises various helix-distorting lesions in DNA and initiates global genome NER. XPC forms a complex with HR23B, XPC-HR23B, and stimulates TDG in the presence of APEX1 [23]. The XPC-HR23B damage-recognition complex is the earliest damage detector in the NER pathway and is essential for the recruitment of all subsequent NER factors, including transcription repair factor TFIIH [24]. Rad23 homolog a (RAD23A) (also known as hHR23A) participates in postreplication repair functions [22]. It is down-regulated by 1.3-fold in AD brain samples. RAD23A is involved in NER through regulation of XPC [25]. Increased levels of RAD23A have been observed in cells that experience oxidative DNA damage.

ROS attack on DNA leads to a variety of modifications of purine and pyrimidine bases that are cytotoxic. BER involves four steps, i.e. removal of a damaged base by a DNA glycosylase, nicking of an apurinic/aprimidinic (AP) site by AP endonuclease, repair synthesis, and finally the sealing of the nick by DNA ligase [26]. BER involves removing the mutated base from the DNA helix and repairing the base with the help of two main enzymes - DNA glycosylases and AP endonucleases. Specific DNA glycosylases recognise different damaged bases. Uracil-dna glycosylase (UNG) and thymine-dna glycosylase (TDG) are DNA glycosylases that recognise uracil and thymine, respectively, in DNA causing its removal [27]. G:U and G:T mispairs in DNA occur by spontaneous deamination of cytosine and 5-methylcytosine [28, 29]. While UNG is up-regulated by 2.7-fold in the AD samples, TDG is down-regulated by 2.8-fold. Current evidence has shown that TDG plays a role in differential gene regulation [30]. Alternative promoter usage and splicing of UNG results in two different isoforms - the mitochondrial UNG1 and the nuclear UNG2 [31]. A recent study states that UNG2 may be the sole enzyme for repair of U:A pairs in human cells and a major glycosylase responsible for repair of G:U mispairs [32]. In the AD samples UNG2 is down-regulated by 1.7-fold.

The BER pathway initiated by DNA glycosylases prevents mutations by removing the oxidative lesions from the DNA [33]. However, BER generates genotoxic intermediates, such as AP sites, which prevent normal DNA replication. Apex nuclease (multifunctional DNA repair enzyme) 1 (APEX1), which encodes the major AP endonuclease in human cells,

repairs oxidative DNA damage in a DNA glycosylase-independent manner [34]. It is a multifunctional enzyme, involved in DNA repair, transcription regulation, oxidative signalling, etc. APEX1 (also known as Ref-1/HAP-1) is involved in the nucleotide incision repair (NIR) pathway and is considered as a back-up to the BER pathway for oxidative DNA base damage [35]. However, since NIR and BER have common substrates, they work in concert to remove potentially mutagenic and cytotoxic lesions in DNA. APEX1 is hypothesised to have a role in protection against cell lethality and suppression of mutations [22]. Loss-of-function studies in mice indicate a critical requirement for APEX1 for animal survival. It has also been suggested that APEX1 is involved in proofreading, removing mismatched nucleotides arising from BER polymerases activity [36]. In the AD samples in this study, APEX1 is decreased by 1.9-fold.

In the AD samples, N-methylpurine-dna glycosylase (MPG), a DNA glycosylase involved in BER, is decreased by 1.6-fold. It is the first enzyme in BER and recognises and removes damaged bases, such as 3-methylpurines, 3-methyladenines and 7-methylguanines, leaving behind an apurinic site that is further processed by other DNA repair proteins [37, 38, 39]. MPG also participates in the repair of 8-hydroxyguanine and hypoxanthine [39]. Cells deficient in MPG have exhibited increased sensitivity to alkylation-induced chromosome damage and cell killing [39].

Polymerase (DNA directed), beta (POLB) is another participant of the BER process [32] and is downregulated by 2.9-fold in the AD individuals. The DNA repair synthesis step in BER is supported by POLB. In the ‘Passing-the-Baton’ model of BER, BER takes place in several steps, i.e. the DNA damaged site is transferred from a DNA glycosylase to APEX1, then from APEX1 to POLB, and finally from POLB to DNA ligase I or III [40]. This depicts how APEX1 and POLB interact in the BER process. Moreover, APEX1 stimulates strand-displacement DNA synthesis by POLB [40]. Single-strand DNA interruptions (SSIs) frequently occur during BER and are rectified by SSI repair subpathways, each mediated by different genes [41]. Recent work suggests that SSIs produced during the repair of alkylated DNA bases are repaired by the pathway mediated by POLB [41]. POLB promotes recruitment of DNA ligase III alpha-XRCC1 to sites of BER and decrease in POLB leads to a reduced presence of DNA ligase III alpha-XRCC1 at the BER sites [42].

Oxidative DNA demethylation is a mechanism for repairing alkylation damage to DNA. AlkB (alkylation repair homologues 5 and 6 (e.coli) (ALKBH5 and ALKBH6)) is a gene involved in this function although its precise function has not yet been resolved. A study has shown that AlkB is particularly essential for the repair of alkylated single-stranded DNA (ssDNA) while another has presented evidence that it is a member of the superfamily of

2-oxoglutarate- and iron-dependent oxygenases (2OG-Fe(II)-oxygenases) [43, 44], thereby suggesting its functions in the cell. ssDNA is transiently present in the human genome at the replication fork and at transcription sites. ssDNA is much more vulnerable to the induction of 1-methyladenine by chemical methylation than double-stranded DNA (dsDNA) [45]. However, since some amount of 1-methyladenine will also be formed by methylation of dsDNA, AlkB may also be necessary for global genome repair [45]. Numerous different AlkB homologues exist in human cells [44]. ALKBH6 is down-regulated by 3.2-fold and ALKBH5 is down-regulated by 1.6-fold in the AD samples of our study.

Fanconi anemia, complementation group g (FANCG) is associated with hypersensitivity to DNA-damaging agents, chromosomal instability, and defective DNA repair [46]. It is a part of the RAD6 pathway which is sensitive towards a variety of genotoxic agents. As a DNA repair protein, it may operate in postreplication repair or in a role of the cell cycle checkpoint guard [22]. It has also been shown to be involved in protection against oxidative DNA damage [46]. FANCG is decreased by 1.3-fold. Fanconi anemia, complementation group f (FANCF) is another gene that is down-regulated by 1.4-fold in AD samples. The function of FANCF is not well documented. FANCF interacts with FANCG and stimulates the assembly of other FA proteins into a stable core complex [47]. FANCA, FANCC, FANCE, FANCF, FANCG, and FANCL assemble in a nuclear core complex, which is essential for the monoubiquitination of FANCD2, which in turn co-localises with BRCA1, RAD51, and PCNA in sites containing other DNA repair proteins [47].

Ribonucleotide reductase m2 b (RRM2B) (also known as p53R2) is reduced by 2-fold in the AD individuals of our study. The family of ribonucleotide reductase (RR) aids in the *de novo* conversion of the ribonucleoside diphosphates to deoxyribonucleoside diphosphates that are important for DNA synthesis and repair [48]. DNA synthesis is an important phase of the DNA repair mechanism and RRM2B assembles deoxyribonucleotides for the synthesis of DNA. P53R2-dependent DNA synthesis is important for cell survival and dysfunction of this pathway can result in activation of p53-dependent apoptosis [49].

Table 1: Number of significant transcripts obtained at the lowest false discovery rates achieved by each dataset using SAM and Limma.

Dataset	SAM results		Limma results	
	Dunckley	1663	FDR: 0.11%	2
Blalock	12	FDR: 4.27%	1414	FDR: 0.49%

Table 2: Some statistically significant biological processes from the set of 1663 significant genes from Dunckley et al.’s data

GO Biological Process	Ease Score
Nervous system development and cellular process	5.93E-08
Homophilic cell adhesion	1.04E-07
Neuron development and cell-cell signaling	3.11E-06
Neurophysiological process and neuron development	2.35E-05
Cell-cell adhesion	7.72E-05
Protein depolymerisation	8.29E-05
Neurogenesis and cell communication	1.91E-04
Neuron maturation and cell differentiation	2.05E-04
Negative regulation of cell organisation and biogenesis and cytoskeleton organisation and biogenesis	3.52E-04
Ion homeostasis and cation transport	0.017
Carbohydrate metabolism and base-excision repair	0.024
DNA repair and carbohydrate metabolism	0.032

## Conclusion

The power of microdissected gene expression data and its advantage over regular dissection gene expression data is evident from this analysis. In both the gene selection methods we adopted in this study, the Dunckley et al.’s data provided a large number of significant genes at very low FDR values. Furthermore, In a disease that has cell loss as one of its most significant phenotype, microarray data obtained via regular dissection cannot unambiguously explain whether a transcript was actually down-regulated in a living cell or was not present due to cell loss. Microarray results validation can be tedious and expensive if measures are not taken to ensure that the transcripts selected for further investigation are not false positives. The aforementioned results indicate that although microdissection is expensive, it provides richer data with less ambiguity in results interpretation. Furthermore, Blalock et al.’s data does not have incipient AD cases as the samples are of Braak stage V AD, and therefore may be an inappropriate dataset to study incipient AD.

AD has been characterised by extensive cell death. Cell death can occur by injury (necrosis) or by suicide (apoptosis). Sometimes cells are induced to commit suicide in order to preserve genomic integrity. Two factors induce a cell to commit suicide - the withdrawal of positive signals, that is, signals needed for continued survival, and the receipt of negative signals. Positive signals include growth factors for neurons, and negative signals include increased levels of oxidants within the cell, DNA damage by these oxidants and the accumulation of misfolded proteins. In AD brains the negative signals clearly overshadow the positive signals. When defenses against DNA damage cannot keep pace with the degree

Table 3: Fifteen DNA repair genes identified from Dunckley et al.’s data. All but two genes are downregulated.

RefSeq ID	Gene name	Fold Change
NM 001641	apex nuclease (multifunctional dna repair enzyme) 1	-1.90
NM 002690	polymerase (DNA directed), beta	-2.85
NM 003211	thymine-dna glycosylase	-2.83
NM 004628	xeroderma pigmentosum, complementation group c	1.46
NM 004629	fanconi anemia, complementation group g	-1.34
NM 005053	rad23 homolog a (s. cerevisiae)	-1.31
NM 005316	general transcription factor IIIH, polypeptide 1, 62kda	-1.45
NM 007254	polynucleotide kinase 3'-phosphatase	-1.95
NM 015713	ribonucleotide reductase m2 b (tp53 inducible)	-2.04
NM 022725	fanconi anemia, complementation group f	-1.42
NM 003362	uracil-dna glycosylase	2.75
NM 001024592	uracil-dna glycosylase 2	-1.68
NM 001015052	n-methylpurine-dna glycosylase	-1.59
NM 017758	alkb, alkylation repair homolog 5 (e. coli)	-1.56
NM 032878	alkb, alkylation repair homolog 6(e. coli)	-3.16

of damage, such as under conditions of high oxidative or alkylating stress, or when repair is inefficient, the miscoding potential of DNA lesions can result in the accumulation of genetic defects that may contribute to early senescence, apoptosis or cancer. Therefore, DNA repair rate is an important determinant of cell pathology [27].

The rate of oxidative DNA damage is elevated in the brain due to the high oxygen consumption rate and the BER pathway is the key player in oxidative DNA damage repair [50]. Furthermore, DNA POLB is hypothesised to be the major player in defective repair in ageing brains [51]. In line with the aforementioned findings, we have noticed the increased presence of genes involved in the BER pathway, many of which have been downregulated in AD samples. Some of the 15 DNA repair genes present in our significant gene list are involved in the early stages of DNA damage detection and some of them help in the recruitment of other downstream proteins to help in DNA repair. In summary, our results indicate that the base excision repair pathway fails to handle the DNA damage repair in an efficient manner and is a compelling explanation for many of the pathological features of LOAD.

## Materials and Methods

### Data

Dunckley et al. dataset consists of 13 normal controls (Braak stages 0–II; average age: 80.1 years) and 20 AD affected (Braak stages III–IV; average age: 84.7 years) samples obtained by laser capture microdissection (LCM) from the entorhinal cortex [7]. Braak stages III–IV is considered ‘incipient’ AD [2, 10]. In Dunckley et al.’s dataset, 1000 neurons were collected from each of the 33 samples via LCM. Blalock et al. dataset consists of 9 normal controls (Braak stage II; average age: 85.3 years) and 22 AD affected (Braak stages V–VI; average age: 85.8 years) samples generated by regular dissection of hippocampal specimens (CA1 and CA3) [6]. The patients were classified as ‘incipient’ based on their MiniMental Status Examination (MMSE) scores. Both datasets consisted of affected samples that showed the NFT pathological phenotype. Moreover, both datasets used Affymetrix GeneChip HGU133A for their gene expression profiling. Each dataset was analysed separately.

### Significant transcripts selection

Both datasets were normalised using gcRMA [52]. Probesets were mapped to genes using DAVID [22]. Probesets that did not map to any gene name and those matching to hypothetical proteins with no known functions, at the time of writing this manuscript, were removed. When multiple probesets mapped to the same gene, only the probeset with the highest mean was selected. This preprocessing resulted in a total of 15827 transcripts for Dunckley et al.’s data and 11955 transcripts for Blalock’s et al.’s data.

In order to compare the power of both datasets in terms of the lowest false discovery rate (FDR) and number of differentially expressed transcripts, we applied two different differentially expressed genes selection methods on the two datasets - SAM [8] and Limma [9]. SAM is an open-source software which uses a modified t-statistics approach to identify differentially expressed genes. Limma is a software package(<http://www.bioconductor.org>) consisting of various linear models to select differentially expressed genes. We used the linear model that uses t-statistics and an empirical Bayes approach to select significant genes.

First SAM and Limma were applied to both datasets, and the FDR and number of significant genes were recorded and results compared. Then the DNA repair genes present in the list of differentially expressed genes obtained from Dunckley et al.’s data was taken

to investigate the DNA repair capacity in incipient AD. DNA repair genes were identified by comparing to the genes obtained with those described in a published article on DNA repair [31] and DAVID [22]. When Gene Ontology (GO) analysis was performed using EASE (<http://niaid.abcc.ncifcrf.gov/home.jsp>) on the complete list of significant genes identified by SAM from Dunckley et al.'s data, the DNA repair category was found to be statistically significant. For the sake of brevity, we did not include the analysis of significant genes obtained from limma.

## Acknowledgement

The authors would like to thank Johannes Freudenberg at the Biomedical Informatics Department in Cincinnati Children's Hospital, OH, for helping with the data normalisation.

## References

- [1] Schneider E (1999) Aging in the third millenium. *Science* 283:796–797.
- [2] Rossler M, Zarski R, Bohl J, Ohm TG (2002) Stage-dependent and sector-specific neuronal loss in hippocampus during alzheimers disease. *Acta Neuropathol* 103:363–369.
- [3] Delacourte A, David JP, Sergeant N, Buee L, Wattez A, et al. (1999) The biochemical pathway of neurofibrillary degeneration in aging and alzheimers disease. *Neurology* 52:1158.
- [4] Gomez IT, Price JL, McKeel DWJ, Morris JC, Growdon JH, et al. (1996) Profound loss of layer II entorhinal cortex neurons occurs in very mild alzheimers disease. *Journal of Neuroscience* 16:4491–4500.
- [5] Emmert-Buck MR, Bonner RF, Smith PD, Chaqui RF, Zhunag Z, et al. (1996) Laser capture microdissection. *Science* 274:998–1001.
- [6] Blalock EM, Geddes JW, Chen KC, Porter NM, Markesbery WR, et al. (2004) Incipient alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc Natl Acad Sci USA* 101:2174–2178.
- [7] Dunckley T, Beach TG, Ramsey KE, Grover A, Mastroeni D, et al. (2006) Gene expression correlates of neurofibrillary tangles in alzheimer's disease. *Neurobiology of Aging* 27(10):1359–1371.
- [8] Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionising radiation response. *Proc Natl Acad Sci USA* 98:5116 – 5121.
- [9] Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* 3(1).

- [10] Braak H, Braak E (1991) Neuropathological staging of alzheimer-related changes. *Acta Neuropathologica* 82:239 – 259.
- [11] Davydov V, Hansen LA, Shackelford D (2003) Is dna repair compromised in alzheimer’s disease? *Neurobiology of Aging* 24:953–968.
- [12] Markesbery WR, Lovell MA (2006) Dna oxidation in alzheimer’s disease. *Antioxid Redox Signal* 8(11-12):2039–2045.
- [13] Gregory RC, Taniguchi T, D’Andrea AD (2003) Regulation of the fanconi anemia pathway by monoubiquitination. *Seminars in Cancer Biology* 13(1):77–82.
- [14] Brooks PJ (2002) Dna repair in neural cells: basic science and clinical implications. *Mutation Research* 509(1-2):93–108.
- [15] Rao KS (2007) Mechanisms of disease: Dna repair defects and neurological disease. *Nature Clinical Practice Neurology* 3:162–172.
- [16] Claire Chappell C, Hanakahi LA, Karimi-Busheri F, Weinfeld M, West SC (2002) Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. *The EMBO Journal* 21:28272832.
- [17] Merlo D, Di Stasi AMM, Bonini P, Mollinari C, Cardinale A, et al. (2005) Dna repair in post-mitotic neurons: a gene-trapping strategy. *Cell Death and Differentiation* 12:307309.
- [18] Karimi-Busheri F, Lee J, Tomkinson AE, Weinfeld M (1998) Repair of dna strand gaps and nicks containing 3’-phosphate and 5’- hydroxyl termini by purified mammalian enzymes. *Nucleic Acids Research* 26(19):4395–4400.
- [19] Meijer M, Karimi-Busheri F, Huang TY, Weinfeld M, Young D (2002) Pnk1, a dna kinase/phosphatase required for normal response to dna damage by gamma -radiation or camptothecin in *schizosaccharomyces pombe*. *Journal of Biological Chemistry* 277(6):4050–4055.
- [20] Lockett KL, Snowwhite IV, Hu JJ (2005) Nucleotide-excision repair and prostate cancer risk. *Cancer Letters* 220(2):125–135.
- [21] Cleaver JE (2000) Common pathways for ultraviolet skin carcinogenesis in the repair and replication defective groups of xeroderma pigmentosum. *J Dermatol Sci* 23:1–11.
- [22] Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, et al. (2003) David: Database for annotation, visualization, and integrated discovery. *Genome Biology* 4:P3.
- [23] Shimizu Y, Iwai S, Hanaoka F, Sugawara K (2003) Xeroderma pigmentosum group c protein interacts physically and functionally with thymine dna glycosylase. *EMBO Journal* 22(1):164173.
- [24] Volker M, Mone MJ, Karmakar P, van Hoffen A, Schul W, et al. (2001) Sequential assembly of the nucleotide excision repair factors in vivo. *Mol Cell* 8(1):213–224.



- [25] Hsieh HC, Hsieh YH, Huang YH, Shen FC, Tsai HN, et al. (2005) Hhr23a, a human homolog of *saccharomyces cerevisiae rad23*, regulates xeroderma pigmentosum c protein and is required for nucleotide excision repair. *Biochem Biophys Res Commun* 335(1):181–187.
- [26] Lindahl T, Wood RD (1999) Quality control by dna repair. *Science* 286(5446):1897 – 1905.
- [27] Wikipedia ([en.wikipedia.org/wiki/dna\\_repair](http://en.wikipedia.org/wiki/dna_repair)) .
- [28] Katafuchi A, Matsuo A, Terato H, Ohyama Y, Ide Y (2003) Repair of oxidative cytosine damage by dna glycosylases. *Nucleic Acids Research Supplement NO 3* :269–270.
- [29] Gallinari P, Jiricny J (1996) A new class of uracil-dna glycosylases related to human thymine-dna glycosylase. *Nature* 383:735 – 738.
- [30] Waters R (2006) Maintaining genome integrity. *EMBO reports* 7(4):377381.
- [31] Wood RD, Mitchell M, Sgouros J, Lindahl T (2001) Human dna repair genes. *Science* 291:1284–1289.
- [32] Akbari M, Otterlei M, Pena-Diaz J, Aas AP, Kavli B, et al. (2004) Repair of u/g and u/a in dna by ung2-associated repair complexes takes place predominantly by short-patch repair both in proliferating and growth-arrested cells. *Nucleic Acids Research* 32(18):54865498.
- [33] Lindahl T (2001) Keynote: past, present and future aspects of base excision repair. *Prog Nucleic Acid Res Mol Biol* 68:xvii–xxx.
- [34] Gros L, Ishchenko AA, Ide H, Elder RH, Sapparbaev MK (2004) The major human ap endonuclease (*ape1*) is involved in the nucleotide incision repair pathway. *Nucleic Acids Research* 32(1):73–81.
- [35] Ischenko AA, Sapparbaev MK (2002) Alternative nucleotide incision repair pathway for oxidative dna damage. *Nature* 415:183187.
- [36] Jiricny J (2002) An ape that proofreads. *Nature* 415(6872):593–594.
- [37] Miao F, Bouziane M, O’Connor TR (1998) Interaction of the recombinant human methylpurine-dna glycosylase (*mpg* protein) with oligodeoxyribonucleotides containing either hypoxanthine or abasic sites. *Nucleic Acids Research* 26(17):4034–4041.
- [38] O Connor TR (1993) Purification and characterization of human 3-methyladenine-dna glycosylase. *Nucleic Acids Research* 21(24):5561–5569.
- [39] Cerda SR, Turk PW, Thor AD, Weitzman SA (1998) Altered expression of the dna repair protein, n-methylpurine-dna glycosylase (*mpg*), in breast cancer. *Federation of European Biochemical Societies Letters* 431(1):12–18.
- [40] Sukhanova MV, Khodyreva SN, Lebedeva NA, Prasad R, Wilson SH, et al. (2005) Human base excision repair enzymes apurinic/apyrimidinic endonuclease1 (*ape1*), dna polymerase ? and poly(adp-ribose) polymerase 1: interplay between strand-displacement dna synthesis and proofreading exonuclease activity. *Nucleic Acids Research* 33(4):1222 – 1229.

- [41] Ho EL, Satoh MS (2003) Repair of single-strand dna interruptions by redundant pathways and its implication in cellular sensitivity to dna-damaging agents. *Nucleic Acids Research* 31(23):7032–7040.
- [42] Parsons JL, Dianova II, Allinson SL, Dianov GL (2005) Dna polymerase beta promotes recruitment of dna ligase iii alpha-xrcc1 to sites of base excision repair. *Biochemistry* 44(31):10613–10619.
- [43] Dinglay S, Trewick SC, Lindahl T, Sedgwick B (2000) Defective processing of methylated single-stranded dna by e. coli alkb mutants. *Genes Dev* 14:2097–2105.
- [44] Aravind L, Koonin EV (2001) The dna-repair protein alkb, egl-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biology* 2:Research 0007.1–0007.8.
- [45] Falnes P, Johansen RF, Seeberg E (2002) Alkb-mediated oxidative demethylation reverses dna damage in escherichia coli. *Nature* 419:178–182.
- [46] Futaki M, Igarashi T, Watanabe S, Kajigaya S, Tatsuguchi A, et al. (2002) The fancg fanconi anemia protein interacts with cyp2e1: possible role in protection against oxidative dna damage. *Carcinogenesis* 23(1):67–72.
- [47] Leveille F, Blom E, Medhurst AL, Bier P, Laghmani EH, et al. (2004) The fanconi anemia gene product fancf is a flexible adaptor protein. *J Biol Chem* 279(38):39421–39430.
- [48] Zhou B, Liu X, Mo X, Xue L, Darwish D, et al. (2003) The human ribonucleotide reductase subunit hrrm2 complements p53r2 in response to uv-induced dna repair in cells with mutant p53. *Cancer Research* 63:6583–6594.
- [49] Yamaguchi T, Matsuda K, Sagiya Y, Iwadate M, Fujino MA, et al. (2001) p53r2-dependent pathway for dna synthesis in a p53-regulated cell cycle checkpoint. *Cancer Research* 61:8256–8262.
- [50] Englander EW, Ma H (2006) Differential modulation of base excision repair activities during brain ontogeny: Implications for repair of transcribed dna. *Mech Ageing Dev* 127(1):64–69.
- [51] Rao KS, Annapurna VV, Raji NS (2001) Dna polymerase-beta may be the main player for defective dna repair in aging rat neurons. *Annals of the New York Academy of Sciences* 928:113–120.
- [52] Irizarry RA, Wu Z, Jaffee HA (2006) Comparison of affymetrix genechip expression measures. *Bioinformatics* 22(7):789–794.