Using induced pluripotent stem cells to explore genetic and epigenetic variation associated with Alzheimer’s disease

Jennifer Imm¹, Talitha L. Kerrigan¹, Aaron Jeffries¹, and Katie Lunnon¹

¹University of Exeter Medical School, Exeter University, Exeter, UK

* Corresponding author: Katie Lunnon, University of Exeter Medical School, RILD Building Level 4, Royal Devon and Exeter Hospital, Barrack Rd, Exeter. EX2 5DW. UK. E-mail: k.lunnon@exeter.ac.uk
ABSTRACT

It is thought that both genetic and epigenetic variation play a role in Alzheimer’s disease initiation and progression. With the advent of somatic cell reprogramming into induced pluripotent stem cells it is now possible to generate patient derived cells that are able to more accurately model and recapitulate disease. Furthermore, by combining this with recent advances in (epi)genome editing technologies it is possible to begin to examine the functional consequence of previously nominated genetic variants and infer epigenetic causality from recently identified epigenetic variants. In this review we explore the role of genetic and epigenetic variation in Alzheimer’s disease and how the functional relevance of nominated loci can be investigated using induced pluripotent stem cells and (epi)genome editing techniques.
Alzheimer's disease

Alzheimer's disease (AD) is the most prevalent neurodegenerative disorder and accounts for approximately 60-80% of all dementia cases worldwide [1]. Dementia is estimated to affect 46.8 million people worldwide, with this set to double every 20 years reaching 131.5 million in 2050 [2]. The disease is characterized by the accumulation of amyloid beta (Aβ) plaques, intracellular neurofibrillary tangles of hyperphosphorylated tau [3] and loss of synaptic connections [4]; taken together these lead to neuronal cell death. This is accompanied by cognitive and behavioral changes, such as memory impairments, language disturbance and hallucinations. The early cognitive decline in AD can be attributed to the degeneration of cholinergic neuronal cells found in the cortical and limbic brain regions such as the hippocampus [5] and the basal forebrain [6].

The deposition of senile plaques and tangles does not occur at random, but follows a distinct and characteristic pattern [7-10], starting in the neocortex and then the hippocampus [11], whilst other regions, such as the cerebellum, remain relatively unaffected [12]. This specific topographical distribution correlates with, and explains, the characteristic symptoms of AD; the hippocampus and neocortex are well known for being involved in controlling emotions, memory and higher brain function [13, 14]. The cerebellum on the other hand is responsible for the coordination, motor and voluntary movements, and there are far fewer aberrations in these in AD patients when compared with the prevalence of other symptoms [15]. There is already considerable pathology before the disease is diagnosed [16], with the onset of symptoms sometimes occurring at least 10 years after Aβ is first deposited [17]. This apparent delay in the appearance of symptoms is caused by there being a threshold of cholinergic loss before the brain can no longer compensate and ameliorate the deficit [11]. Although much progress has been made in understanding the cellular pathology of AD, the
treatments currently available only temporarily alleviate some symptoms and do not modify the underlying pathology.

**Genetic variation associated with AD**

Given the high heritability estimates (~60-80%) for AD based on quantitative genetic studies [18], initial etiological studies have focussed on identifying a genetic basis for disease. Although some AD cases are caused by autosomal dominant mutations in three genes (APP, PSEN1, PSEN2), these account for less than 5% of AD prevalence and are early-onset, occurring before the age of 65 years. Most AD cases are late-onset (>65 years) and sporadic, with no defined etiology. However in recent years, large cohort collections and the relatively inexpensive cost of assessing genetic variation through genome-wide association studies (GWAS) has allowed the identification of common variants associated with risk of developing AD. These studies have demonstrated that late onset Alzheimer's disease (LOAD) is thought to be multifactorial with many different genes and single nucleotide polymorphisms (SNPs) being implicated in, and contributing to, disease onset and progression [18]. The most robustly associated gene with LOAD is Apolipoprotein E (APOE), which encodes a polymorphic glycoprotein that is involved in the transport of cholesterol and other lipids [19] alongside neuronal growth [20] and tissue repair [21]. There are three isoforms of APOE that all correspond to allelic variation at a single locus, ε2, ε3 and ε4, which can be distinguished by cysteine to arginine substitutions at the amino acid positions 112 and 158 [22]. The ε4 variant confers increased risk of developing LOAD, with each additional copy of the risk allele lowering the mean age of onset [23]. Whilst APOE ε4 accounts for approximately 20% of genetic risk for developing LOAD it cannot explain all of disease incidence, as not everyone who is homozygous for ε4 actually develops AD [24]. Aside from APOE there are numerous other risk loci (SNPs) that have been implicated in AD. The most recent meta-analysis of nearly 75,000 individuals nominated 19 common genetic variants, of which 11 were novel disease loci [25] (Table 1). Interestingly, many of the GWAS loci that have been nominated for AD can be linked to amyloid processing or
inflammation. Whilst risk variants that have been identified from GWAS only confer a relatively modest effect size, with odds ratios (ORs) between 0.73 and 1.22 per loci investigated [26], it is thought that these could act cumulatively to cause the onset of degeneration. Scientists have generated polygenic risk scores (PRS) for AD, which combine the effects of many disease-associated SNPs to predict disease risk [27] and recently it has been reported that the PRS prediction captures nearly all common genetic risk for AD [28]. However, another study has demonstrated that collectively common SNPs for AD only account for a third of phenotypic variance in AD [29]. Recent efforts to explain the missing heritability of AD have used sequencing approaches to identify rare variants, with a larger effect size, with SNPs in PLD3, TREM2, TM2D3 and PICALM being nominated in recent years [30-34].

**Table 1: Summary table of SNPs associated with AD, which reach genome wide significance in Lambert et al, 2013 [25].** Abbreviations: Chr. = Chromosome

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr:Position</th>
<th>Closest Gene</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6656401</td>
<td>1:207692049</td>
<td>CR1</td>
<td>1.18</td>
</tr>
<tr>
<td>rs6733839</td>
<td>2:127892810</td>
<td>BIN1</td>
<td>1.22</td>
</tr>
<tr>
<td>rs10948363</td>
<td>6:47487762</td>
<td>CD2AP</td>
<td>1.10</td>
</tr>
<tr>
<td>rs11771145</td>
<td>7:143110762</td>
<td>EPHA1</td>
<td>0.90</td>
</tr>
<tr>
<td>rs9331896</td>
<td>8:27467686</td>
<td>CLU</td>
<td>0.86</td>
</tr>
<tr>
<td>rs983392</td>
<td>11:59923508</td>
<td>MS4A6A</td>
<td>0.90</td>
</tr>
<tr>
<td>rs10792832</td>
<td>11:85867875</td>
<td>PICALM</td>
<td>0.87</td>
</tr>
<tr>
<td>rs4147929</td>
<td>19:1063443</td>
<td>ABCA7</td>
<td>1.15</td>
</tr>
<tr>
<td>rs3865444</td>
<td>19:51727962</td>
<td>CD33</td>
<td>0.94</td>
</tr>
<tr>
<td>rs9271192</td>
<td>6:32578530</td>
<td>HLA-DRB5–HLA-DRB1</td>
<td>1.11</td>
</tr>
<tr>
<td>rs28834970</td>
<td>8:27195121</td>
<td>PTK2B</td>
<td>1.10</td>
</tr>
<tr>
<td>rs11218343</td>
<td>11:121435587</td>
<td>SORL1</td>
<td>0.77</td>
</tr>
<tr>
<td>rs10498633</td>
<td>14:92926952</td>
<td>SLC24A4 RIN3</td>
<td>0.91</td>
</tr>
<tr>
<td>rs8093731</td>
<td>18:29088958</td>
<td>DSG2</td>
<td>0.73</td>
</tr>
<tr>
<td>rs35349669</td>
<td>2:234068476</td>
<td>INPP5D</td>
<td>1.08</td>
</tr>
<tr>
<td>rs190982</td>
<td>5:88223420</td>
<td>MEF2C</td>
<td>0.93</td>
</tr>
<tr>
<td>rs2718058</td>
<td>7:37841534</td>
<td>NME8</td>
<td>0.93</td>
</tr>
<tr>
<td>rs1476679</td>
<td>7:100004446</td>
<td>ZCWPW1</td>
<td>0.91</td>
</tr>
<tr>
<td>rs10838725</td>
<td>11:47557871</td>
<td>CELF1</td>
<td>1.08</td>
</tr>
<tr>
<td>rs17125944</td>
<td>14:53400629</td>
<td>FERMT2</td>
<td>1.14</td>
</tr>
<tr>
<td>rs7274581</td>
<td>20:55018260</td>
<td>CASS4</td>
<td>0.88</td>
</tr>
</tbody>
</table>
Despite the fact that GWAS and sequencing efforts have been successful in identifying novel genes involved in AD, the majority of SNPs lie outside of coding regions of the genome. These variants are thus unlikely to have direct structural or functional effects on their gene’s protein product, and are more likely to affect gene regulation at other loci. By integrating genetic variation with transcriptomic measurements in the same samples, for example from microarray or RNA sequencing experiments, it is possible to correlate disease associated genetic variants with changes in gene expression to identify expression quantitative trait loci (eQTLs). Such disease-associated eQTLs can occur both within the same gene (in cis) or distally within another gene (in trans). Recent research indicates that SNPs may change expression by initially altering the binding ability of one ‘pioneering’ transcription factor (TF), which then recruits other TFs. This was shown to occur at a CTCF motif, which if disrupted affects the binding of five different TFs [35]. However, if there are two or more variants in perfect linkage disequilibrium (LD) with one another, the non-random association of loci at different genomic locations, then it is not possible to distinguish which variant is acting as the eQTL. To further complicate matters, many eQTLs act in a cell type specific manner [36]. This is particularly relevant when studying heterogeneous tissues such as the brain or blood or in a disease such as AD, when cellular abundance is known to be altered [37, 38]. In the context of AD, one recent study demonstrated an enrichment for monocyte-specific eQTLs at disease-associated loci, suggesting a role of the innate immune system in AD pathology [39]. In support of this, Karch et al, tested whether GWAS LOAD SNPs act as cis-eQTLs for LOAD GWAS genes [40]. They were able to show the AD SNP rs1476679 in ZCWPW1, was significantly associated with the expression of PILRB and GATS in most brain regions, including the hippocampus. Interestingly, PILRB acts as a binding partner for TYROBP, both of which can be found on microglia [41]. The expression of TYROBP is restricted to cells of the immune system such as microglia, and has been shown to be up-regulated in the brains of patients with LOAD. In a recent gene-regulatory network analysis TYROBP was shown to be a key causal regulator of a microglial/immune module highlighted as the module most associated with pathophysiology of LOAD. This microglia/immune module was also shown to
contain a number of AD GWAS risk loci, such as CD33, MS4A4A and MS4A6A [39]. Furthermore, TREM2, which has been recently nominated from sequencing studies, is known to interact and signal through TRYOBP [39]. Taken together many network level analyses have highlighted a role for microglia and neuroinflammation in AD risk [39, 40, 42, 43].

**A role for epigenomic dysfunction in AD?**

Recently, increased understanding about the functional complexity of the genome has led to growing recognition about the likely role of non-sequence-based "epigenetic" variation in AD [44]. Epigenetic processes mediate the reversible regulation of gene expression, occurring independently of DNA sequence, acting principally through chemical modifications to DNA and nucleosomal histone proteins (Figure 1). The most widely studied epigenetic modification in human studies is that of DNA methylation, given it is the most stable and longest lasting change.

In general, *in vitro* studies have suggested that there is a trend towards global DNA hypomethylation in AD. For example, one study demonstrated global hypomethylation in a glioblastoma cell line with the APP<sup>SWE</sup> mutation, which occurs in familial AD [45]. Likewise, it was shown that there are lower DNA methylation levels in brain microvascular endothelial cells that were exposed to high levels of synthetic Aβ<sub>1-40</sub> [46]. However, in contrast to this a more recent study using neuroblastoma cells exposed to synthetic Aβ<sub>1-40</sub> showed no significant change in DNA methylation levels [47]. In the context of human post-mortem brain samples there are also some conflicting results; contrasting studies have shown decreased levels of global DNA methylation in AD-associated brain regions such as the entorhinal cortex [48] and hippocampus [49], whilst others report no change [50, 51] or even increased levels in the frontal cortex [52] and hippocampus [53].
Figure 1. Diagram to illustrate the different epigenetic mechanisms that have been identified. (A) Regulation of chromatic structure through post-translational modifications to histone proteins. This can include: acetylation, methylation, SUMOylation, ubiquitylation, citrullination and ADP-riboseylation. (B) Addition of chemical tags to DNA to the 5’ end of a cytosine nucleotide. This creates 5-methylcytosine (5mC – the most commonly studied epigenetic mark), 5-hydroxymethylcytosine (5hmC), 5-carboxylcytosine (5caC) and 5-formylcytosine (5fC). (C) Small RNA molecules, such as microRNA can also affect gene expression either through degrading mRNA or altering protein translation.

In recent years, advances in genomic technology have allowed the first genome-scale studies of DNA methylation in AD [54, 55]. To date, all published studies have utilized the Illumina Infinium 450K Methylation Beadarray (450K array) to examine DNA methylation changes at >485,000 loci in post-mortem brain samples. Although these epigenome-wide association studies (EWAS) have been performed on independent sample cohorts, in a range of anatomically distinct cortical brain regions, a number of consistently differentially methylated genes have been identified [56]. Most notably one such gene to be robustly hypermethylated in AD cortex is Ankyrin1 (ANK1) [54, 55]. Interestingly, disease-associated ANK1 hypermethylation has been observed in a tissue-specific manner, with brain regions
affected by AD pathology (entorhinal cortex, superior temporal gyrus and prefrontal cortex) showing significant hypermethylation, whilst the cerebellum, a region largely unaffected by pathology, and pre-mortem blood show no disease-associated changes. The majority of genes identified by EWAS are distinct from those nominated in GWAS with the exception of BIN1. However, despite different genes being identified with the two approaches many of these do reside in common pathways [57]. Since the publication of the initial EWAS, there have been additional studies identifying additional differentially methylated genes in AD, including CRTC1 [58], APOE [59] and TREM2 [60].

To date the majority of AD EWAS have focussed on DNA methylation, with histone modifications considerably less well studied. However, one study has shown that the histone deacetylase (HDAC) inhibitor, sodium butyrate, can improve cognitive function, synaptic density and plasticity in mouse models with inducible overexpression of p25 protein [61]. P25 has been shown to be linked to numerous features of AD such as amyloid and tau pathology alongside memory loss and neurodegeneration [62-64]. These findings have been corroborated by another, who also demonstrated improved memory function after treatment with sodium butyrate in transgenic APPPS1-21 mice [65]. Furthermore, memory improvements were seen even when it was administered at the latest stages of amyloid pathology and were associated with increased expression of genes implicated in associative learning. Other studies have also shown HDACs to have therapeutic utility, such as Depakote [66], phenylbutyric acid [67] and trichostatin A [68]. Whilst these studies are encouraging and show the potential utility of HDACs as therapeutic agents there are, however, a couple of considerations with this approach. First, most models, including the ones we describe, are models of familial AD, which use small numbers of animals and, as such, results should not be over interpreted. Alongside this, these studies use pan-HDAC inhibitors, which are all known to affect multiple HDACs. Whilst it is useful to know that there may be some involvement of histone modifications or HDACs in AD initiation or progression,
these studies do not highlight specifically which ones are implicated. In order to address this, a number of groups have crossed APPPS1-21 mice with those lacking certain HDACs. APPPS1-21 mice that lack HDAC5 show exaggerated memory impairment [69], whilst those lacking HDAC6 have improved memory, but this is achieved without changing amyloid deposition [70]. Another HDAC that could have potential as a therapeutic target is HDAC2, as mice with reduced HDAC2 show increased memory and synaptic plasticity [71]. Therefore, it appears that only specific HDACs show promise as potential drug targets. When using certain HDACs it is important to be aware that they can also affect DNA methylation. For instance it has been shown that Depakote can cause extensive DNA methylation changes including demethylation changes at specific genes such as MMP2, MAGEB2 and WIF1, which have been implicated in tumour growth and metastasis [72]. This study supports the concept that various epigenetic marks are in a dynamic relationship with one another and that you cannot necessarily target one without affecting another. Finally, as with any potential drug that is tested on model systems, such as mice, it is important to remember that these rarely translate effectively into humans. One good example of this comes from studies which demonstrated the utility of anti-amyloid antibodies [73]. In these APP-transgenic mice, just one injection of m266, an anti-APP mouse antibody, was able to reverse cognitive deficits without reducing amyloid plaque burden [73]. However, when taken to phase three clinical trials, the human antibody, Solanezumab, did not reduce cognitive decline in those with mild dementia due to AD [74].

**Induced pluripotent stem cells: new models for late-onset Alzheimer's disease?**

To fully understand and elucidate the mechanisms of disease etiology, extensive modelling must take place. Traditionally, this has been achieved by a number of methods, including both animal (murine) models and primary patient cell lines. Whilst both of these approaches have their own merits, they can prove inconvenient and do not completely and accurately reflect human disease. At present, the AD research field has had a heavy focus on disease modelling through the use of transgenic mouse models [75], as there is a well-developed
understanding of genetic manipulation techniques in this organism. Furthermore, mice are more phylogenetically related to humans than other simpler model organisms such as *drosophila melanogaster* and *caenorhabditis elegans*, although these do allow for more experimental control than mice. Due to the close relation of mice and humans they also have good utility for studying familial AD, through the use of transgenic mice containing mutations in the *APP* and *PSEN* genes. This has led to advances in our understanding of multiple aspects of AD, in particular amyloid pathology and the differential effects of the various Aβ peptides. However, despite the extensive use of these transgenic models to study AD, they do not accurately recapitulate AD, as the mice do not display overt neurodegeneration [76-78] or have amyloid plaques [79]. Whilst proven useful for modelling autosomal disease, such as familial AD, the mouse models do not have extensive utility for studying sporadic AD, which has both polygenic and environmental components. Even if it were possible to model the genetics of sporadic AD in transgenic mice the effect sizes of each associated variant would be small and therefore difficult to determine phenotypic outcome. However, there have been murine studies that have targeted replacement of the endogenous murine *Apoe* gene with human *APOE-ε4*. These mice demonstrated reduced spatial learning and a reduction in dendritic spine density in the medial entorhinal cortex [80]. In another study where *APOE* (both ε3/ε3 and ε4/ε4) mice were crossed with mice containing a mutant human form of *APP*, the *APOE-ε4xAPP* mice displayed significantly worse spatial memory performance than their *APOE-ε3xAPP* counterparts, but this was also associated with insulin dysfunction [81].

A more promising avenue for modelling SNPs in complex diseases, such as sporadic AD, is through the use of stem cell technology. Embryonic stem cells (ESCs), which are derived from the inner cell mass of an embryo (blastocyst), have the ability to differentiate into any cell in the body [82]. Due to their inherent plasticity, and as genomic variation can be assessed relatively inexpensively through PCR, microarray, or sequencing technology, there is the potential that they could be used to study the effect of disease-associated SNPs on
the functionality of specific cell types. However, whilst useful, the ethical issues implicated with using embryo derived ESCs are numerous. Recent advances in stem cell technology have allowed the production of stem cells derived from adult tissue, such as blood, urine and keratinocytes [83]. These induced pluripotent stem cells (iPSCs) have almost identical characteristics to ESCs: they share the same morphology, can differentiate into any cell type in the body, have unlimited growth and have the same expression pattern of genes [84]; potentially making them a very powerful tool in research.

There are, however, a number of caveats when utilizing iPSCs to model complex diseases that must be considered. Associated with inducing pluripotency are the global cellular epigenetic changes that allow the cells to alter gene expression in order for them to be functionally identical to ESCs. Despite being functionally identical, several groups report that iPSCs have different DNA methylation profiles and gene expression patterns to ESCs [85-88]. Some groups attribute this variation due to an ‘epigenetic memory’ where iPSCs show residual DNA methylation patterns that are typical of the tissue they originate from [89]. These differentially methylated regions (DMRs) were shown to affect the differentiation potential of the newly formed iPSCs. For example, iPSCs derived from neural and fibroblast progenitors maintained DNA methylation marks at sites associated with haematopoietic lineages, which decreased the potential for these iPSCs to form blood cells. Subsequently, it is possible to reverse these restricting methyl marks by increasing the cells passage number or treatment with chromatin modifying compounds [89]. This treatment is associated with a decrease in DNA methylation at haematopoietic loci and therefore an increase in blood cell fate potential. Therefore, although it would appear that this epigenetic memory can affect the differentiation potential of cells initially, this effect is actually only transient. It has also been observed that certain subsets of cells can become stuck in a partially reprogrammed state. This is due to inefficient DNA demethylation at certain sites or the incomplete repression of TFs [90]. Despite this, these aberrations can be rectified using RNA inhibition of TFs or treatment with DNA methylase inhibitors. Another potential source of epigenetic variation
between ESCs and iPSCs is the microenvironment in which the iPSCs have been generated. Cooper and Newman have demonstrated that there is some correlation between cells’ gene expression patterns and the laboratory the cell lines are derived from [91, 92]. This demonstrates that the environment can affect the epigenome and therefore downstream gene expression of cell lines. To fully assess the differences in the epigenomes between iPSCs and ESCs, Lister et al, utilized a shotgun bisulphite sequencing technique (MethylC-seq) to look at the whole-genome DNA methylome at single base-pair resolution [93]. This demonstrated that, overall, ESCs and iPSCs are similar, but that there are some inherent differences between their DNA methylomes. The reprogramming of somatic cells generated hundreds of DMRs that could be attributed to both memory from the somatic cell and iPSC specific DNA methylation patterns that are susceptible during the reprogramming process as many DMRs were consistent across independent iPSC lines [86]. All of these studies demonstrate that there are fundamental differences in both the epigenome and gene expression patterns of ESCs and iPSCs. However, there are ways to rectify some of these differences meaning that iPSCs still have utility as disease models, although the differences must be taken into account when interpreting results.

Since the introduction of iPSC technology there have been increasingly more studies utilizing iPSCs for disease modeling and small molecule testing as, theoretically, iPSCs are an exact genetic match of the patient they are derived from. To accurately model AD using iPSCs, the generation of specific neuronal populations are usually necessary, particularly iPSC derived cortical and cholinergic neurons. Generating these requires certain factors, including SB431542 and LDN-193189 [94], which act as inhibitors of TGF-β I and BMP type I receptors respectively. This inhibition prevents SMAD phosphorylation, supressing cellular renewal and promoting cortical differentiation [95, 96]. A good iPSC model of AD would not only be the correct cell type of interest, but also show the neuropathological features and characteristics of the disease. There have been several studies that have reported that iPSCs show certain disease features [97-100]. More specifically, in AD iPSCs have also
been used to show Aβ-induced synaptotoxicity [101]. In this study, Nieweg et al demonstrated not only that Aβ altered AMPA receptors post-synaptically and impaired axonal vesicle clustering, but also increased the phosphorylation status of tau, another key characteristic hallmark of AD. Alongside aberrations in tau phosphorylation, degeneration of cortical neurons is also a very prominent clinical feature of AD, and it is believed that this causes the onset of symptoms. Therefore, to truly understand AD as a disease, being able to recapitulate this neuronal cell death is vital. One recent study has shown that iPSC derived basal forebrain cholinergic neurons (BCFNs) heterozygous for APOE (ε3/ε4) are more susceptible to glutamate mediated cell death, whilst also showing an increased Aβ42/Aβ40 ratio when compared to BCFNs generated from healthy age matched control patients [102].

One of the major utilities for using iPSCs to study AD is the ability to examine the effects of genetic variants with a relatively small effect size on phenotype. One study which takes advantage of this has demonstrated that iPSC-derived neurons carrying genetic variants in SORL1, which increase LOAD risk, have reduced response to BDNF treatment. This not only manifests at the level of SORL1 expression but also impacts APP processing [103]. Furthermore, given that AD has a polygenic component, it is also possible to assess the effect of different combinations of disease-associated SNPs. As it is possible to use iPSCs to generate patient specific neuronal cells, there is the potential to generate libraries of cells with varying combinations of LOAD-associated SNPs and therefore different susceptibilities to disease. Interestingly, two studies have demonstrated this variability in susceptibility using LOAD iPSC-derived neurons [104, 105]. In the first, undertaken by Israel et al, they found that neurons from one patient, but not from another, showed AD-associated phenotypes. This included altered levels of secreted Aβ1-40, higher aGSK3 levels and had significantly increased p-tau/total tau [104]. The second study, conducted by Kondo et al, showed differential intracellular Aβ oligomer accumulation, inducing endoplasmic reticulum and oxidative stress [105]. Taken together these studies show how different genetic backgrounds
can alter disease initiation and progression as well as the complex genetic interplay there is in LOAD. However, being able to investigate the effects of different polygenic risk scores on living and developing neuronal populations will provide more valuable insight into the role genetic variants play in terms of physiological/cellular aberrations and disease progression in LOAD. Importantly, data that is generated through the usage of iPSCs can therefore be compared and contrasted with data collected from molecular studies in human post-mortem tissue to potentially elucidate the disease-specific effects.

**Using genetic editing to elucidate the functional consequence of disease-associated variation**

Using iPSCs and recent advances in (epi)genetic editing technology it is becoming possible to start teasing apart the underlying mechanisms that may be driving AD pathogenesis. Clustered regularly interspersed short palindromic repeats (CRISPR) uses RNA guided Cas9 nuclease to introduce DNA breaks which can be repaired through homologous recombination, indel mutations or with a vector carrying a desired mutation [106]. CRIPSR can be used provided that the sequence of interest is unique compared to the rest of the genome and is upstream of a protospacer adjacent motif (PAM) sequence. The PAM sequence is typically three to five nucleotides long and serves as a binding region for the Cas9 to bind. Unfortunately, this is a requirement of the method and can be technically challenging. Despite the PAM sequences being relatively common throughout the human genome they can often be in the incorrect location relative to the sequence of interest and can make modifying the gene difficult. Furthermore, if the target locus has high homology to another region in the genome then there is the potential for off-target effects resulting in inadvertent mutations [107, 108].

The main advantage of the CRISPR-Cas9 system is that you can create isogenic control lines that only show genetic variation at your disease-specific loci. One recent example of this in the AD field comes from Pires et al, who successfully corrected the A79V PSEN1
mutation in a patient AD iPSC cell line [109]. These types of control lines are extremely beneficial for studying disease-associated genetic variation, as they enable the minimization of genetic variability as both disease and control lines have the same genetic background. Such isogenic lines have been recently used to study familial AD [105, 110, 111]. One study has shown iPSCs harbouring the APPSWE and PSEN1 M146V mutations have increased total Aβ production, and up to a threefold increase in the Aβ42/Aβ40 ratio when compared to their isogenic controls. These changes have been shown to correlate to neuronal identity, maturity and mutation load [112].

Whilst it proves relatively simple to use this technique to study the effects of causative mutations, such as those in familial AD, other genetic variants like SNPs associated with sporadic AD are more problematic to model. This is due to the fact that they may only be relevant to diseases such as AD in specific combinations. As an extension to the CRISPR system, it is possible to alter multiple loci using CRISPR-multiplexing [113]. In this system multiple guide RNAs are assembled into the same vector and transfected into the cells allowing the targeting of multiple loci. This tool will undoubtedly prove incredibly useful for LOAD research as it will allow researchers to modify up to seven loci. By modifying multiple disease-associated SNPs, one could investigate the effects of various combinations on cell physiology, protein expression and aggregation. This approach would work very well for the LOAD risk SNPs BIN1, CLU and PICALM, for example, as these have all been shown to interact with Aβ/tau [114-117]. Therefore by altering these specific SNPs one could investigate how tau and Aβ pathology changes over time. Another interesting point is that this methodology would allow one to investigate the interaction between disease-associated SNPs with reported relationships, for example PICALM and the APOE ε4 allele. In a previous study of familial AD patients, a homozygous PICALM genotype (rs3851179) was shown to modulate prefrontal cortex volume and cognitive impairment in carriers of the APOE ε4 allele [118]. As both proteins are involved in the same Aβ clearance pathway [119, 120], it is thought that alterations in the endocytic functions of PICALM may synergistically
affect APOE ε4. This could mean there is a higher likelihood of Aβ remaining in the brain and therefore increasing plaque formation.

Establishing Causality

An important step for research is to establish whether disease-associated variation is causing disease. This is simpler to test when examining genetic variation, as we know that the SNP has been present in an individual throughout their life course, prior to disease onset. However, when investigating the functional effects of disease-associated epigenetic variation, the relationship is less clear, and it is difficult to determine whether epigenetic changes are a cause, or a consequence of, the disease process. Whilst previous post-mortem brain studies have provided valuable physiologically relevant information about epigenetic changes occurring at later disease stages it is difficult to establish if those alterations actually initiated disease, therefore making it problematic to infer causality. In any disease it is critical to be certain of what is causing disease, but particularly so in AD. First because this is crucial for the design of effective drug therapies to combat disease, and second, because AD neuropathology generally occurs at least 10 years before symptomatic onset, it is important to understand the root cause(s) to be able to monitor and diagnose the disease in its very earliest stages. iPSCs have real utility for determining whether associations identified in EWAS analyses could represent a secondary effect of disease, or could be driving disease progression. Using a modified version of the CRISPR-Cas9 technology it has become possible to alter DNA methylation at specific loci. By fusing the Cas9 protein with the enzymatic domains of TET1 or DNMT3A it is possible to remove, or add methyl groups to DNA respectively [121]. This method builds on previous work [122, 123] where Tet1 and Dnmt3a were fused to TALE proteins to achieve the same effect. By being able to manipulate the epigenetic landscape of cells, particularly at loci associated with disease it will enable researchers to establish whether disease-associated epigenetic variation is causative. These techniques have already been utilized in diseases such as cancer [124, 125], but to our knowledge have not yet been reported in AD research. The
epigenetic status of genes such as *ANK1* which have robust changes in AD would be an excellent initial target for this new methodology. It has the potential to reveal whether the DNA methylation changes seen are truly causative and precede disease initiation, or whether they are a consequence of progression. Even if it is the latter and it is found that these alterations do not cause disease, but occur in the earliest stages of the disease, they could potentially be used as a biomarker for diagnosis prior to symptomatic onset. Alongside this, such changes also tell us something about the disease process and its progression. Given that recent studies have demonstrated robust epigenetic alterations in AD, iPSCs could become a valuable tool in which these studies could be taken further. However, to undertake epigenetic studies in these cells it will be very important to fully understand the epigenetic landscape of the iPSCs themselves, both throughout differentiation and at maturity.

**Current challenges to Progress**

Whilst it is possible to create disease relevant cells using iPSCs, as with any model, iPSCs also have their limitations. A large concern when using iPSC-derived neurons to model disease is that the resulting neurons represent an immature and fetal population [126-129]. This is of particular concern when studying diseases of aging, such as LOAD. However, efforts are being made to overcome this particular issue. For example by expressing a mutant form of *LMNA*, which is known to cause premature aging. By expressing progerin in the iPSC-derived dopaminergic neurons it was possible to create phenotypes that were resultant of both the induced aging and genetic susceptibility [130]. The aged Parkinson’s disease neurons had marked dendrite neurodegeneration, reduced tyrosine-hydroxylase expression and displayed epigenetic markers of aging that were not present in the control populations. Another study directly reprogrammed fibroblasts into neurons, skipping iPSC pluripotency stages, in an attempt to overcome iPSC immaturity [131]. As this protocol does not induce pluripotency the inducible neurons (iNs) display both age-related epigenetic and transcriptomic signatures showing age-associated decreases in RanBP17, a nuclear
transport receptor. Whilst these neurons would prove useful for studying diseases of aged cells, such as AD, it is difficult to make large amounts of primary material as these cells cannot be propagated, unlike the iPSC cells.

**FUTURE PERSPECTIVE**

Since the pioneering work by Takahashi and Yamanaka in 2006, the use of iPSCs has been able to significantly advance complex disease research. They have enabled researchers to more accurately recapitulate disease phenotypes in a cell culture system. Whilst iPSCs are far from being used therapeutically, they have proven useful for investigating the molecular and genetic underpinnings of LOAD. Once there have been more extensive investigations into the effects of SNP burden and their molecular targets, iPSCs can be used to test the effectiveness of new therapeutic interventions. Although at present iPSC generation and differentiation are costly and time consuming, differentiation protocols are quickening, and the use of an individual's own iPSCs to select their appropriate treatment would be a first step towards personalized medicine, potentially improving the patient’s life.

Whilst it is still unknown whether the global epigenetic changes that occur during iPSC generation affect the end epitype of cells, there is still promise that these cells could be used to study the epigenetics of complex diseases. If epigenetic aberrations do prove to be an issue, then these will have to be taken into account during experimental design and analysis. However, before identified changes can be targeted for therapeutic intervention it will be important to determine whether they are causal; with the recent advances in genetic and epigenetic editing technology this will soon be possible. Finally, while there are many questions that still remain unanswered and many challenges ahead when addressing these, with the correct model and methodologies these will hopefully be overcome.
EXECUTIVE SUMMARY

Alzheimer's disease (AD)

- Most prevalent neurodegenerative disorder causing 60-80% of dementia cases worldwide
- Characterised by extracellular depositions of amyloid beta (Aβ) protein and intracellular neurofibrillary tangles of paired helical filaments of tau

Genetic variation associated with AD

- Through the use of genome wide association studies (GWAS) a number of single nucleotide polymorphisms (SNPs) have been associated with late onset AD
- Many GWAS loci implicated in disease have been linked to amyloid processing and inflammation

A role for epigenomic dysfunction in AD?

- Recent epigenome wide association studies (EWAS) have identified a number of loci that are differentially methylated in disease
- The majority of genes identified by EWAS are distinct from those nominated in GWAS with the exception of BIN1.

Induced pluripotent stem cells (iPSCs): new models for late-onset AD?

- Through somatic cell reprogramming it is possible to generate induced pluripotent stem cell (iPSC) derived neuronal cells
- These iPSC-derived neuronal cells have been shown to reflect some disease features
- iPSC-derived neuronal cells can be used to assess the effect of polygenic risk on physiological/cellular changes and disease progression

Using genetic editing to elucidate the functional consequence of disease-associated variation

- Clustered regularly interspersed short palindromic repeats (CRISPR) uses RNA guided Cas9 nucleases to introduce modifications in the genome
• CRISPR-multiplexing can be used to edit multiple loci within the genome

**Establishing Causality**

• iPSC models have utility in determining whether loci identified from GWAS and EWAS are causative in the disease process

• Using CRISPR the epigenetic landscape of cells can be altered to establish whether DNA methylation changes associated with disease are causative

**FINANCIAL DISCLOSURES/ ACKNOWLEDGEMENTS**

This work was funded by an Alzheimer’s Society project grant to KL (grant number AS-PG-2 14-038), an Alzheimer’s Research UK network co-operation grant to KL (grant number ARUK-NCG2017A-5) and an Alzheimer’s Association new investigator research grant to KL (grant number NIRG-14-320878). JI is supported by the Alzheimer’s Society Doctoral Training Centre in Dementia Research at the University of Exeter (grant number AS-DTC-2014-030) and the Garfield Weston Foundation. The authors declare no conflicts of interest.

**REFERENCES AND REFERENCE ANNOTATIONS**

1. Association AS. What Is Dementia?


**Identified 19 loci associated with AD, 11 of which had not previously been implicated, highlighting new pathways that could play a role in AD.


31. Jakobsdottir J, Van Der Lee SJ. Rare Functional Variant in TM2D3 is Associated with Late-Onset Alzheimer's Disease. 12(10), e1006327 (2016).


36. Zhernakova DV, Deelen P. Identification of context-dependent expression quantitative trait loci in whole blood. doi:10.1038/ng.3737 (2016).


** Used postmortem brain tissue to construct a gene-regulatory network of LOAD. Demonstrated immune and microglia specific network which contained TYROBP, a binding partner of TREM2, as a key regulator.


** One of first EWAS published and robustly implicated epigenetic dysregulation of ANK1 in AD.


** One of first EWAS published


** Demonstrated which factors were important for the maintenance and induction of the stem cell phenotype, therefore allowing for the generation of iPSCs.


** Found a robust and reproducible way of creating cortical neurons from iPSCs, a method for investigating AD with iPSCs.


** Demonstrates that iPSC-derived neurons are able to recapitulate disease phenotypes and are more susceptible to glutamate meditated cell death.


** Demonstrate that iPSC-derived neurons do not retain age related signatures, but induced neurons do display age-specific transcriptmic signatures and have age associated decreases in RanBP17.