An Alzheimer's disease-associated common regulatory variant in PTK2B has causal effects on microglial function

Erica Bello[1,2,3*], Kathleen Long[1], Sho Iwama[1], Juliette Steer[1,2], Sarah Cooper[1,2], Kaur Alasoo[1,4], Natsuhiko Kumasaka[1,5], Jeremy Schwartzentruber[1,2], Nikolaos I Panousis[1], Andrew Bassett[1,2,*]

[1] Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK
[2] OpenTargets, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK
[3] Milner Therapeutics Institute, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge, CB2 0AW, UK
[4] Institute of Computer Science, University of Tartu, Tartu, Estonia
[5] National Center for Child Health and Development, Tokyo, Japan

*To whom correspondence should be addressed. Tel: +44(0)1223 494933; Email: ab42@sanger.ac.uk, eb956@cam.ac.uk

Abstract

Genome-wide association studies (GWAS) are revealing an ever growing number of genetic associations with disease, but identifying and functionally validating the causal variants underlying these associations is very challenging and has only been done for a vanishingly small number of variants. Here we validate a causal single nucleotide polymorphism (SNP) associated with an increased risk of Alzheimer's disease (AD) in an intronic enhancer of the *PTK2B* gene, by engineering it into human induced pluripotent stem cells (hiPSCs). Upon differentiation to macrophages and microglia, this variant shows effects on chromatin accessibility of the enhancer and gene expression changes of the PTK2B gene, and results in global changes to the transcriptome and phenotype of these cells. Expression of interferon gamma responsive genes including chemokine transcripts and their protein products are altered, and chemotaxis of the resulting microglial cells is affected. This variant thus causes disease-relevant transcriptomic and phenotypic changes, and we propose that it acts by altering microglia reactivity, consistent with the role of these cells in progression of AD.

Introduction

Alzheimer's disease (AD) is the most common form of dementia, and has a strong genetic component estimated to be 60-80%[1]. AD is characterised by the presence of β-amyloid-containing plaques (Aβ) and intracellular neurofibrillary tangles of hyperphosphorylated tau, accompanied by neuroinflammation, neuronal, and synapse loss[2]. Multiple GWAS studies have collectively identified more than 80 genetic loci associated with disease[3–7], the vast majority of which do not affect protein coding sequence. Due to linkage disequilibrium at these loci, it is difficult to identify the causal variants. Multiple computational methods have been developed to prioritise variants that cause disease including statistical fine mapping, co-localisation with expression or chromatin accessibility quantitative trait loci (eQTL, caQTL), overlap with enhancer regions identified by e.g. assay for transposase accessible chromatin (ATAC) or transcription factor binding in disease relevant cell types. However, none of these methods are able to unambiguously assign causal variants and, given the vanishingly small number of truly validated variants, it is difficult to benchmark their effectiveness. Experimental methods have also been developed to identify or prioritise variants such as massively parallel reporter assays (MPRA[8]) and genome engineering-based interrogation of enhancers (GenIE[9], GenIE-ATAC[10]) but these are limited to understanding the effects of variants on transcription or chromatin accessibility and do not directly demonstrate the downstream effects of these polymorphisms. Genome engineering of isogenic human cell models, such as differentiated hiPSCs differing in a single variant, provide an attractive means for understanding the role of specific SNPs on cell function and have been successful in understanding the role of strong effect familial mutations[11,12,13]. However, such systems have yet to be successfully applied to more than a handful of common genetic variants[14,15,16] and their effects are often confounded by variation between edited clones and during differentiation[17].

Neuroinflammation has been shown to be a crucial factor in many neurodegenerative disorders, especially AD[18]. Microglia are resident immune cells of the central nervous system and perform various functions under both homeostatic and disease conditions [19]. Microglia play important roles in synaptic remodelling, synaptic pruning and phagocytosis of dead neurons[20]. Microglia have been implicated in AD by human genetics studies because of specific expression of AD-related genes in these cells[21–23], colocalisation of microglial eQTLs with AD GWAS hits[24,25] and implication of rare causal variants in microglial-specific genes, such as TREM2[26] and CD33[26,27]. Nevertheless, the exact role of microglia in AD pathogenesis is still unclear. Aβ and tau aggregates are able to activate microglia and elicit an inflammatory response that can have two functions: it can be neuroprotective by increasing Aβ or tau clearance, but it can also contribute to Aβ and tau production and spreading and induce neurodegeneration and synaptic loss[28,29]. Microglia are known to be involved in clearance of aggregated proteins such as Aβ[30] and imbalanced microglial activation or insufficient microglial phagocytic capacity can accelerate the

accumulation of extracellular Aβ[31]. However, depletion of microglia in amyloid precursor protein-transgenic mouse models was found to have little effect on plaque formation or maintenance. Microglia have also been found to potentially contribute to tauopathy by spreading tau protein via exosomes across brain regions[32] as well as its uptake and degradation[29]. Moreover, sustained microglia activation appears to contribute to neurodegeneration by impacting synaptic plasticity and causing neuronal damage[33,34]. As well as microglia, other myeloid cell types, such as infiltrating monocytes[35,36], choroid plexus[37] and perivascular macrophages[38–40], have also been found to play a role in AD pathogenesis.

PTK2B is a non-receptor tyrosine kinase that is expressed in neurons, microglia, astrocytes, monocytes and tissue-resident macrophages, and has diverse physiological and pathological roles including cell adhesion, cell migration, inflammatory responses, tumor invasiveness[41], neuronal development and plasticity[42,43]. PTK2B is a member of the focal adhesion kinase (FAK) subfamily and after autophosphorylation it recruits Src-family kinases[44] and interacts with multiple partners to activate various downstream signalling pathways, such as the MAPK/ERK pathway[42]. In neurons, PTK2B can be activated following stimulation of glutamate receptors[45], neuronal depolarization[46] and increased intracellular free Ca2+[47]. In neurons PTK2B is implicated in synaptic plasticity[48] and neurite outgrowth[49]. In non-neuronal cells, such as osteoclasts[50], macrophages[51–53] and monocytes[54], PTK2B is involved in cell migration and regulation of actin cytoskeleton downstream of integrins. In monocytes, the ROS-sensitive calcium channel TRPM2 can activate the PTK2B/ERK pathway, leading to nuclear translocation of nuclear factor-kB and increased chemokine production[55]. This pathway is activated by Aβ in microglia, where PTK2B is autophosphorylated and generates a positive feedback loop to further activate TRPM2[56]. PTK2B has been implicated in AD through a number of GWAS studies[4,6,57–64]. As well as being involved in response to Aβ, PTK2B co-localizes with hyperphosphorylated Tau in the brain of AD patients and in a mouse model, and it may contribute to Tau phosphorylation[65,66]. Evidence suggests that genetic deletion of Ptk2b rescues a number of Aβ-associated phenotypes, such as memory impairment, synapse loss and impaired synaptic plasticity, in APP/PS1 animal models of AD[67]. Data collected from hemizygous PS19 (MAPT P301S, 1N4R) transgenic mice crossed with Ptk2b$^{-/-}$ animals, points to a protective role for Ptk2b against Tau phosphorylation and Tau-induced behavioral deficits, in part through the suppression of LKB1 and p38 MAPK[68].
Several lines of evidence point to PTK2B as an important player in the pathophysiology of AD but much remains to be learnt about the exact role of PTK2B in AD pathogenesis.

Results

Prioritisation of a putative Alzheimer's disease causal variant at the PTK2B locus

GWAS have identified two independent genetic associations at the PTK2B-CLU locus[57]. The signal over the PTK2B gene colocalises with both an eQTL and caQTL in hiPSC-derived

116 macrophages (Fig 1a)[69]. Re-analysis of publicly available datasets[25,69,70] shows that the risk
117 haplotype is linked to both a downregulation of PTK2B gene expression in hiPSC-derived
118 macrophages (Fig 1b and 1c) and an increase in chromatin accessibility at an enhancer
119 within intron 6 of the PTK2B gene (Fig 1d). Interestingly, in this case there is a single variant
120 in the credible set (rs28834970) that lies within the ATAC peak that is changed in
121 accessibility in the caQTL analysis, strongly suggesting that this variant is causal (C being
122 the risk allele). Note that there is some inconsistency in published literature on the
123 directionality of these effects[69,25,70], but this data has been reanalysed in conjunction with
124 the authors, and in additional hiPSC-derived macrophage datasets so we are confident of
125 the directionality of these changes described here. This confirmed that the C allele of
126 rs28834970 decreases expression of *PTK2B* in hiPSC-derived macrophages (Fig 1b and Fig
127 1c, whereas in monocytes and one study of primary microglia, the C allele of rs28834970
128 is associated with increased expression of *PTK2B* (Supp Fig 1a)[69].
129
130 **Characterisation of rs28834970 in hiPSC-derived macrophages and microglia**
131 In a recent study this variant was found to be associated with an increase in PTK2B
132 expression only in blood of AD patients and not in any brain region examined[71]. Given the
133 important role of PTK2B in myeloid cells and the evidence suggesting this variant modifies
134 expression of PTK2B in this cell type, we decided to investigate the role of this variant in
135 macrophages and microglia. We edited the C allele of rs28834970 into a well characterised
136 hiPSC line (KOLF2_C1) that was homozygous for the T allele at this position (and thus the
137 non-risk haplotype) using CRISPR-mediated homology directed repair[72] (Suppl Fig 1b and
138 Methods). This allows us to unambiguously study the role of rs28834970 independently
139 of the other variants in linkage disequilibrium. We subsequently analysed three
140 independent homozygous (C/C) clones and two unedited (T/T) clones, one of which had
141 been through the editing process, and the second being the parental line. These were
142 differentiated in three biological replicates into either macrophages[73,74] or microglia[75,76]
143 followed by a number of assays including transcriptomics (RNAseq), chromatin
144 accessibility (ATACseq) and quantitative digital RT-PCR (dd-qRT-PCR) to measure gene
145 expression (Fig 1e). Differentiation was highly reproducible between the two different
146 genotypes by FACS for macrophage markers CD14, CD16 and CD206 (Supp Fig 2a) and
147 microglia markers CD11b and CD45 (Supp Fig 2b). This was further validated by analysis of
148 macrophage and microglial marker gene expression which were mostly consistent with the
149 differentiation trajectory and did not cluster by genotype (Supp Fig 2c). Similarly, there did
150 not appear to be genotype-dependent change in expression of canonical markers of
151 macrophage or microglial activation (Supp Fig 2d).
152
153 **rs28834970 reduces PTK2B expression by creating a novel CEBPB binding site in hiPSC-**
154 **derived macrophages and microglia**
155 We then analysed the effect of the rs28834970 allele on *PTK2B* expression and function of
156 the intronic enhancer element. Consistent with the prediction from the eQTL analysis in

157 hiPSC-derived macrophages, both RNAseq and digital droplet qPCR (dd-qPCR) analysis
158 showed that expression of *PTK2B* was modestly reduced by introduction of the C allele in
159 macrophages (RNAseq: p=0.019 and logFC=-0.2, ddqPCR: p=0.0045 and logFC=-0.39) (Fig
160 2a) and microglia (RNAseq: p=0.06 logFC=-0.57, ddqPCR: p=0.22 and logFC=-0.71) (Fig 2b).
161 The effect was consistent between RNAseq and dd-qPCR, and across microglia and
162 macrophages, but was variable in magnitude. This is consistent with the small effect size
163 in the eQTL analysis (Fig 1b and Fig 1c), and is likely a feature of the majority of common
164 variants that would be expected to have only small or context-specific effects on target
165 gene expression[16].
166
167 Given that an enhancer can have long range effects and mapping of chromosomal
168 interactions at this locus in microglia had shown that enhancers harboring AD-risk variants
169 interacted with active promoters of not only PTK2B and CLU, but also TRIM35 and
170 CHRNA2[77], we also analysed expression of genes in a 600 kb window around rs28834970
171 (Supp Fig 3a). The effect on PTK2B expression was the most consistent change in microglia
172 and macrophages in this window, but there was additionally a slight upregulation of CLU
173 expression specifically in macrophages (logFC=0.72 p-value=0.03) which was not observed
174 in microglia (Supp Fig 3b). This implies that in macrophages, there may be an effect of the
175 rs28834970 enhancer on CLU expression as well as PTK2B. We also analysed changes in
176 splicing of the PTK2B gene and saw only small changes between the two genotypes (Supp
177 Fig 3c), none of which affected splicing of the exons near to the variant.
178
179 We next analysed the effect of rs28834970 on chromatin accessibility, which showed that
180 the C allele created a novel region of accessible chromatin in both macrophages and
181 microglia (Fig 2c) which was consistent with the expectations from the caQTL analysis in
182 hiPSC-derived macrophages (Fig 1d). Analysis of chromatin accessibility within a 600 kb
183 window around the SNP showed that there was no change in most peaks, and the only
184 statistically significant change consistent between macrophages and microglia was the
185 new peak over rs28834970 (macrophages: p=1.65e-16 logFC=2.71, microglia: p=4.19e-12
186 logFC=2.77) (Supp Fig 3d). Peaks within the CHRNA2 and CLU genes were significantly
187 reduced in microglia, and an intergenic peak was increased in macrophages, but these
188 were not consistent between the two cell types, or with the expression changes of these
189 genes (Supp Fig 3d and Supplementary Table 1).
190
191 Further analysis of the sequence around rs28834970 showed that the C allele increased
192 the probability of binding of transcription factors CEBPA and CEBPB (Fig 2d). We thus
193 analysed CEBPB binding to chromatin by CUT&RUN, which showed an increased binding
194 for the C allele of rs28834970 (p=0.04 logFC=2.09)(Fig 2c), consistent with CEBPB binding
195 being responsible for the change in chromatin accessibility. Indeed, in a genome-wide
196 analysis, this was the only region that showed a significant change in CEBPb binding
197 between the C/C and T/T alleles at rs28834970 (Fig 2e)(Supplementary Table 2). CEBPA

198 and CEBPB have previously been shown to have both activator and repressor functions[78],
199 and thus we propose that this novel binding event represses expression of *PTK2B*.
200
201 Taken together, these results validate that rs28834970 within the PTK2B locus is the causal
202 variant and acts through introducing a novel CEBPB binding site that increases chromatin
203 accessibility at the intronic enhancer, and results in repression of PTK2B expression.
204
205 **rs28834970 causes strong effects on the transcriptome including chemokine expression**
206 We next analysed the effects of editing rs28834970 on genome-wide gene expression in
207 macrophages and microglia. Principal component analysis showed that although there was
208 a strong clone and batch effect, the two genotypes were well separated by the second
209 principal component (Fig 3a). Similar results were seen with analysis of global chromatin
210 accessibility, where genotypes were separated by the second principal component (Supp
211 Fig 4a). Analysis of the differentially expressed genes between genotypes showed 690
212 downregulated and 428 upregulated genes in macrophages, and 760 downregulated and
213 328 upregulated in microglia (Fig 3a) (Supplementary Table 3). Importantly, there was a
214 significant ($p<0.0001$) overlap in the differentially expressed genes between the two cell
215 types (Supp Fig 4b), further corroborating that we were identifying true effects of the
216 genotype, not simply variation between edited clones or differentiation artefacts.
217
218 We performed gene set enrichment analysis of the differentially expressed genes relative
219 to all genes expressed in the respective cell type. This showed that a number of gene
220 ontology pathways relevant to microglial or macrophage function were enriched in the
221 downregulated gene set including chemokine-mediated signalling, cellular response to
222 interferon gamma (IFNγ) and migration and chemotaxis (Fig 3b, Fig 3c). These pathways
223 were frequently overlapping between microglia and macrophages (Supp Fig 4c), again
224 supporting that these changes were genotype-dependent. Conversely, there were no
225 obviously relevant gene ontology (GO) enrichments in the upregulated gene set (Supp Fig
226 4d) which may be due to the smaller number of genes analysed.
227
228 Further analysis of the downregulated genes confirmed that there was considerable
229 overlap between microglia and macrophages across the different GO categories (Fig 3d).
230 Of note, we found that the Macrophage Receptor With Collagenous Structure (*MARCO*)
231 gene was among the top 3 downregulated genes in both microglia and macrophages with
232 the C/C allele (Supp Fig 4e). Given their role in microglial and macrophage function, we
233 focused on chemokines, MHC complex proteins and surface receptors and validated their
234 expression by dd-qPCR. All of the chosen genes showed some level of downregulation as
235 expected in microglia, and all except one (HLA-DQA1) were downregulated in
236 macrophages (Fig 3e), with effect sizes generally consistent between dd-qPCR and RNAseq.
237
238

239 **rs28834970 causes changes in chemokine release and migration**
240 Guided by the consistent effects on chemokine and migration-related gene expression in
241 both macrophages and microglia, we next analysed the release of 13 chemokines using a
242 multiplexed assay on microglial cell supernatants. In unstimulated cells, we saw significant
243 reduction in production of 10 of the 13 chemokines tested with the C/C allele at
244 rs28834970, consistent with the results of the transcriptomic analysis (Fig 4a). CCL22 was
245 unexpectedly upregulated, and CXCL2 and CXCL5 were unchanged (Fig 4a).
246
247 Given that several of these differentially expressed chemokines are produced by microglia
248 in response to stimuli such as IFNγ and lipopolysaccharide (LPS), we also analysed whether
249 their production was altered in microglia harbouring the C/C allele in the presence of these
250 stimuli. As expected, both treatments increased production of most (10 of 12) of the
251 chemokines (Supp Fig 5a and Supp Fig 5b). CCL1 production was increased by LPS
252 stimulation in microglia harbouring the C/C allele but not in microglia with the T/T allele
253 (Supp Fig 5b). Similar to what was observed in unstimulated cells, production of most of
254 the chemokines was reduced in the C/C allele at rs28834970 when microglia were
255 stimulated with IFNγ, but the magnitude of the reduction was greatly reduced, such that
256 only 5 of the 12 chemokines were significant (Fig 4b). A similar response was seen upon
257 stimulation with LPS (Supp Fig 5c). Taken together, these data suggest that the C/C allele
258 results in a lower basal chemokine release, and in the presence of stimuli such as IFNγ, this
259 reduction becomes less marked.
260
261 We further investigated the effect of rs28834970 on microglial migration and chemotaxis
262 using a transwell assay in the presence or absence of IFNγ and with or without a
263 chemoattractant relevant to microglia, complement 5a (C5a). Microglia harbouring the
264 unedited (T/T) allele showed chemotaxis towards C5a that was inhibited by IFNγ treatment
265 (Fig 4c). Microglia harbouring the C/C alleles at rs28834970 showed no significant
266 chemotaxis towards C5a or inhibition of migration following IFNγ treatment (Fig 4c). Taken
267 together, these data suggest that microglia harbouring the C/C allele are less sensitive to
268 the chemoattractant C5a and to the inhibitory effect of IFNγ on migration.
269
270 In summary, we show that the C allele of rs28834970 lowers chemokine release by
271 microglia, reduces chemotaxis towards C5a and alters their response to IFNγ stimulation,
272 perhaps by altering microglia reactivity.
273
274 ## Discussion
275 One of the major challenges in interpreting GWAS is to understand the genes and causal
276 variants underlying the genetic association with disease. The use of isogenic cell lines
277 generated by CRISPR in model systems such as hiPSC differentiated cell types offer one
278 way to dissect the role of individual variants in cellular phenotypes independently of other
279 variants in linkage disequilibrium. However, common regulatory variants that are

280 identified in GWAS have rarely been linked to changes in cellular 'omics or phenotypes.
281 Here we identify a likely causal variant from comparison of Alzheimer's disease GWAS,
282 eQTL and caQTL studies and engineer this using CRISPR/Cas9 into hiPSCs to generate
283 isogenic cell lines differing in the single variant. We show that the C/C allele of rs28834970
284 likely introduces a novel CEBPB binding site in an intron of PTK2B that results in
285 transcriptional repression of PTK2B expression. Even though the downregulation of PTK2B
286 caused by the variant does not reach statistical significance in our models, it is consistent
287 with previous reports that this variant causes a decrease in PTK2B expression in
288 monocytes[79] and macrophages[25,69,70] (Fig 1b and c). This could be due to the small effect
289 size making it difficult to study with a limited number of isogenic cell clones or
290 compensatory regulatory mechanisms that may buffer expression levels. Nevertheless, we
291 found that the C/C allele at rs28834970 caused a large transcriptional response and
292 resulting phenotypic changes, including reduction in chemokine release and decreased
293 chemotaxis towards C5a. Downregulation of genes involved in cellular response to IFNγ
294 and the lack of inhibition of migration following IFNγ treatment, are phenotypes consistent
295 with the downregulation of PTK2B in cells harbouring the C/C allele. Following IFNγ
296 receptor stimulation, PTK2B has been shown to bind to JAK2 and contribute to the
297 activation of STAT1 and MAPK pathways[80]. Moreover, inhibition of PTK2B has been shown
298 to decrease GSK-3β-mediated IFNγ signalling[80]. We propose that these phenotypic
299 changes are the result of a change in microglia reactivity, causing them to respond
300 differently to stimulations such as IFNγ, and this may impact the ability of the microglia to
301 respond to aggregated proteins such as Aβ plaques and cellular damage in the brain of
302 Alzheimer's disease patients. Indeed several pro-inflammatory chemokines, such as CCL2,
303 CCL5 and CXCL1, which are downregulated in microglia with the C/C allele both in naive
304 state and when stimulate with IFNγ, are produced by microglia in response to Aβ and are
305 important for Aβ-induced microglia chemotaxis[81–83]. We believe this study demonstrates
306 the power of isogenic cell lines to identify and validate individual variants, including
307 common regulatory SNPs identified from GWAS.
308

309 ## References

310 1. Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch. Gen.*
311 *Psychiatry* **63**, (2006).
312 2. Knopman, D. S. *et al.* Alzheimer disease. *Nature Reviews Disease Primers* **7**, 1–21 (2021).
313 3. The complex genetic architecture of Alzheimer's disease: novel insights and future directions.
314 *eBioMedicine* **90**, 104511 (2023).
315 4. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related
316 dementias. *Nat. Genet.* **54**, (2022).
317 5. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk
318 loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
319 6. Schwartzentruber, J. *et al.* Genome-wide meta-analysis, fine-mapping and integrative
320 prioritization implicate new Alzheimer's disease risk genes. *Nat. Genet.* **53**, (2021).

7. Ghoussaini, M. *et al.* Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2021).
8. Klein, J. C. *et al.* A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nat. Methods* **17**, 1083–1091 (2020).
9. Cooper, S. E., Schwartzentruber, J., Bello, E., Coomber, E. L. & Bassett, A. R. Screening for functional transcriptional and splicing regulatory variants with GenIE. *Nucleic Acids Res.* **48**, e131 (2020).
10. Cooper, S., Schwartzentruber, J., Coomber, E. L., Wu, Q. & Bassett, A. Screening for functional regulatory variants in open chromatin using GenIE-ATAC. *Nucleic Acids Res.* (2023) doi:10.1093/nar/gkad332.
11. Kwart, D. *et al.* A Large Panel of Isogenic APP and PSEN1 Mutant Human iPSC Neurons Reveals Shared Endosomal Abnormalities Mediated by APP β-CTFs, Not Aβ. *Neuron* **104**, (2019).
12. Penney, J., Ralvenius, W. T. & Tsai, L.-H. Modeling Alzheimer's disease with iPSC-derived brain cells. *Mol. Psychiatry* **25**, 148–167 (2019).
13. Moore, S. *et al.* APP Metabolism Regulates Tau Proteostasis in Human Cerebral Cortex Neurons. *Cell Rep.* **11**, 689–696 (2015).
14. APOE4 Causes Widespread Molecular and Cellular Alterations Associated with Alzheimer's Disease Phenotypes in Human iPSC-Derived Brain Cell Types. *Neuron* **98**, 1141–1154.e7 (2018).
15. Hall-Roberts, H. *et al.* TREM2 Alzheimer's variant R47H causes similar transcriptional dysregulation to knockout, yet only subtle functional phenotypes in human iPSC-derived macrophages. *Alzheimers. Res. Ther.* **12**, 151 (2020).
16. Soldner, F. *et al.* Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. *Nature* **533**, 95–99 (2016).
17. Volpato, V. *et al.* Reproducibility of Molecular Phenotypes after Long-Term Differentiation to Human iPSC-Derived Neurons: A Multi-Site Omics Study. *Stem Cell Reports* **11**, 897–911 (2018).
18. Labzin, L. I., Heneka, M. T. & Latz, E. Innate Immunity and Neurodegeneration. *Annu. Rev. Med.* **69**, (2018).
19. Wang, H. Microglia Heterogeneity in Alzheimer's Disease: Insights From Single-Cell Technologies. *Front. Synaptic Neurosci.* **13**, 773590 (2021).
20. Li, Q. & Barres, B. A. Microglia and macrophages in brain homeostasis and disease. *Nat. Rev. Immunol.* **18**, 225–242 (2017).
21. Li, Y. *et al.* Genomics of Alzheimer's disease implicates the innate and adaptive immune systems. *Cell. Mol. Life Sci.* **78**, 7397–7426 (2021).
22. Nott, A. & Holtman, I. R. Genetic insights into immune mechanisms of Alzheimer's and Parkinson's disease. *Front. Immunol.* **14**, 1168539 (2023).
23. Tansey, K. E., Cameron, D. & Hill, M. J. Genetic risk for Alzheimer's disease is concentrated in specific macrophage and microglial transcriptional networks. *Genome Med.* **10**, (2018).
24. Lopes, K. P. *et al.* Genetic analysis of the human microglial transcriptome across brain regions, aging and disease pathologies. *Nat. Genet.* **54**, (2022).
25. Young, A. M. H. *et al.* A map of transcriptional heterogeneity and regulatory variation in human microglia. *Nat. Genet.* **53**, (2021).
26. Ulland, T. K. & Colonna, M. TREM2 — a key player in microglial biology and Alzheimer disease. *Nat. Rev. Neurol.* **14**, 667–675 (2018).
27. Zhao, L. CD33 in Alzheimer's Disease – Biology, Pathogenesis, and Therapeutics: A Mini-

366 Review. *Gerontology* **65**, 323–331 (2018).
367 28. Novoa, C. *et al.* Inflammation context in Alzheimer's disease, a relationship intricate to define.
368 *Biol. Res.* **55**, 1–18 (2022).
369 29. Vogels, T., Murgoci, A.-N. & Hromádka, T. Intersection of pathological tau and microglia at the
370 synapse. *Acta Neuropathologica Communications* **7**, 1–25 (2019).
371 30. Rogers, J., Strohmeyer, R., Kovelowski, C. J. & Li, R. Microglia and inflammatory mechanisms
372 in the clearance of amyloid β peptide. *Glia* **40**, 260–269 (2002).
373 31. Hansen, D. V., Hanson, J. E. & Sheng, M. Microglia in Alzheimer's disease. *J. Cell Biol.* **217**,
374 (2018).
375 32. Asai, H. *et al.* Depletion of microglia and inhibition of exosome synthesis halt tau propagation.
376 *Nat. Neurosci.* **18**, (2015).
377 33. Hong, S. *et al.* Complement and microglia mediate early synapse loss in Alzheimer mouse
378 models. *Science* **352**, (2016).
379 34. Miao, J. *et al.* Microglia in Alzheimer's disease: pathogenesis, mechanisms, and therapeutic
380 potentials. *Front. Aging Neurosci.* **15**, 1201982 (2023).
381 35. Monoranu, C. M. *et al.* Is There Any Evidence of Monocytes Involvement in Alzheimer's
382 Disease? A Pilot Study on Human Postmortem Brain. *Journal of Alzheimer's disease reports* **5**,
383 (2021).
384 36. Malm, T., Koistinaho, M., Muona, A., Magga, J. & Koistinaho, J. The role and therapeutic
385 potential of monocytic cells in Alzheimer's disease. *Glia* **58**, 889–900 (2010).
386 37. Gião, T., Teixeira, T., Almeida, M. R. & Cardoso, I. Choroid Plexus in Alzheimer's Disease-The
387 Current State of Knowledge. *Biomedicines* **10**, (2022).
388 38. Park, L. *et al.* Brain Perivascular Macrophages Initiate the Neurovascular Dysfunction of
389 Alzheimer Aβ Peptides. *Circ. Res.* **121**, (2017).
390 39. De Schepper, S. *et al.* Perivascular cells induce microglial phagocytic states and synaptic
391 engulfment via SPP1 in mouse models of Alzheimer's disease. *Nat. Neurosci.* **26**, (2023).
392 40. Munawara, U. *et al.* Hyperactivation of monocytes and macrophages in MCI patients
393 contributes to the progression of Alzheimer's disease. *Immun. Ageing* **18**, 1–25 (2021).
394 41. Zhu, X., Bao, Y., Guo, Y. & Yang, W. Proline-Rich Protein Tyrosine Kinase 2 in Inflammation and
395 Cancer. *Cancers* **10**, (2018).
396 42. Beck, T. N., Nicolas, E., Kopp, M. C. & Golemis, E. A. Adaptors for disorders of the brain? The
397 cancer signaling proteins NEDD9, CASS4, and PTK2B in Alzheimer's disease. *Oncoscience* **1**, 486
398 (2014).
399 43. de Pins, B., Mendes, T., Giralt, A. & Girault, J.-A. The Non-receptor Tyrosine Kinase Pyk2 in
400 Brain Function and Neurological and Psychiatric Diseases. *Front. Synaptic Neurosci.* **13**, 749001
401 (2021).
402 44. Day, R. N., Day, K. H. & Pavalko, F. M. Direct visualization by FRET-FLIM of a putative
403 mechanosome complex involving Src, Pyk2 and MBD2 in living MLO-Y4 cells. *PLoS One* **16**, (2021).
404 45. Bartos, J. A. *et al.* Postsynaptic clustering and activation of Pyk2 by PSD-95. *J. Neurosci.* **30**,
405 (2010).
406 46. Siciliano, J. C., Toutant, M., Derkinderen, P., Sasaki, T. & Girault, J. A. Differential regulation of
407 proline-rich tyrosine kinase 2/cell adhesion kinase beta (PYK2/CAKbeta) and pp125(FAK) by
408 glutamate and depolarization in rat hippocampus. *J. Biol. Chem.* **271**, (1996).
409 47. Lev, S. *et al.* Protein tyrosine kinase PYK2 involved in Ca(2+)-induced regulation of ion channel
410 and MAP kinase functions. *Nature* **376**, (1995).

411 48. Girault, J. A., Costa, A., Derkinderen, P., Studler, J. M. & Toutant, M. FAK and PYK2/CAKbeta in
412 the nervous system: a link between neuronal activity, plasticity and survival? *Trends Neurosci.* **22**,
413 (1999).
414 49. Ivankovic-Dikic, I., Grönroos, E., Blaukat, A., Barth, B. U. & Dikic, I. Pyk2 and FAK regulate
415 neurite outgrowth induced by growth factors and integrins. *Nat. Cell Biol.* **2**, (2000).
416 50. Kong, L. *et al.* Integrin-associated molecules and signalling cross talking in osteoclast
417 cytoskeleton regulation. *J. Cell. Mol. Med.* **24**, (2020).
418 51. Miao, L., Xin, X., Xin, H., Shen, X. & Zhu, Y.-Z. Hydrogen Sulfide Recruits Macrophage Migration
419 by Integrin β1-Src-FAK/Pyk2-Rac Pathway in Myocardial Infarction. *Sci. Rep.* **6**, 1–10 (2016).
420 52. Okigaki, M. *et al.* Pyk2 regulates multiple signaling events crucial for macrophage morphology
421 and migration. *Proc. Natl. Acad. Sci. U. S. A.* **100**, (2003).
422 53. Du, Q.-S. *et al.* Inhibition of PYK2-induced actin cytoskeleton reorganization, PYK2
423 autophosphorylation and focal adhesion targeting by FAK. *J. Cell Sci.* **114**, 2977–2987 (2001).
424 54. Das, P. *et al.* A PKCβ-LYN-PYK2 Signaling Axis Is Critical for MCP-1-Dependent Migration and
425 Adhesion of Monocytes. *J. Immunol.* **206**, (2021).
426 55. Yamamoto, S. *et al.* TRPM2-mediated Ca2+ influx induces chemokine production in
427 monocytes that aggravates inflammatory neutrophil infiltration. *Nat. Med.* **14**, 738 (2008).
428 56. Alawieyah, S. M. S., Sim, J. A., Neubrand, V. E. & Jiang, L. H. A critical role of TRPM2 channel
429 in Aβ42 -induced microglial activation and generation of tumor necrosis factor-α. *Glia* **66**, (2018).
430 57. Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for
431 Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
432 58. Beecham, G. W. *et al.* Genome-wide association meta-analysis of neuropathologic features of
433 Alzheimer's disease and related dementias. *PLoS Genet.* **10**, (2014).
434 59. Jiao, B. *et al.* Polygenic Analysis of Late-Onset Alzheimer's Disease from Mainland China. *PLoS*
435 *One* **10**, (2015).
436 60. Li, Y. Q. *et al.* Common variant in PTK2B is associated with late-onset Alzheimer's disease: A
437 replication study and meta-analyses. *Neurosci. Lett.* **621**, (2016).
438 61. Wang, X. *et al.* Genetic Determinants of Survival in Patientswith Alzheimer's Disease. *J.*
439 *Alzheimers. Dis.* **45**, (2015).
440 62. Nettiksimmons, J., Tranah, G., Evans, D. S., Yokoyama, J. S. & Yaffe, K. Gene-based aggregate
441 SNP associations between candidate AD genes and cognitive decline. *Age* **38**, (2016).
442 63. Lawingco, T. *et al.* Genetic variants in glutamate-, Aβ-, and tau-related pathways determine
443 polygenic risk for Alzheimer's disease. *Neurobiol. Aging* **101**, (2021).
444 64. Rajabli, F. *et al.* Multi-ancestry genome-wide meta-analysis of 56,241 individuals identifies
445 LRRC4C, LHX5-AS1 and nominates ancestry-specific loci PTPRK , GRB14 , and KIAA0825 as novel risk
446 loci for Alzheimer's disease: the Alzheimer's Disease Genetics Consortium. *medRxiv : the preprint*
447 *server for health sciences* (2023) doi:10.1101/2023.07.06.23292311.
448 65. Dourlen, P. *et al.* Functional screening of Alzheimer risk loci identifies PTK2B as an in vivo
449 modulator and early marker of Tau pathology. *Mol. Psychiatry* **22**, (2017).
450 66. Köhler, C., Dinekov, M. & Götz, J. Active glycogen synthase kinase-3 and tau pathology-related
451 tyrosine phosphorylation in pR5 human tau transgenic mice. *Neurobiol. Aging* **34**, (2013).
452 67. Salazar, S. V. *et al.* Alzheimer's Disease Risk Factor Pyk2 Mediates Amyloid-β-Induced Synaptic
453 Dysfunction and Loss. *J. Neurosci.* **39**, (2019).
454 68. Brody, A. H. *et al.* Alzheimer risk gene product Pyk2 suppresses tau phosphorylation and
455 phenotypic effects of tauopathy. *Mol. Neurodegener.* **17**, 1–33 (2022).

456 69. Alasoo, K. *et al.* Shared genetic effects on chromatin and gene expression indicate a role for
457 enhancer priming in immune response. *Nat. Genet.* **50**, 424–431 (2018).
458 70. Panousis, N. I. *et al.* Gene expression QTL mapping in stimulated iPSC-derived macrophages
459 provides insights into common complex diseases. *bioRxiv* 2023.05.29.542425 (2023)
460 doi:10.1101/2023.05.29.542425.
461 71. Tan, M.-S. *et al.* Associations of Alzheimer's disease risk variants with gene expression,
462 amyloidosis, tauopathy, and neurodegeneration. *Alzheimers. Res. Ther.* **13**, (2021).
463 72. Bruntraeger, M., Byrne, M., Long, K. & Bassett, A. R. Editing the Genome of Human Induced
464 Pluripotent Stem Cells Using CRISPR/Cas9 Ribonucleoprotein Complexes. *Methods Mol. Biol.* **1961**,
465 153–183 (2019).
466 73. van Wilgenburg, B., Browne, C., Vowles, J. & Cowley, S. A. Efficient, long term production of
467 monocyte-derived macrophages from human pluripotent stem cells under partly-defined and fully-
468 defined conditions. *PLoS One* **8**, e71098 (2013).
469 74. Douthwaite, H., Arteagabeitia, A. B. & Mukhopadhyay, S. Differentiation of Human Induced
470 Pluripotent Stem Cell into Macrophages. *Bio Protoc* **12**, e4361 (2022).
471 75. Brownjohn, P. W. *et al.* Functional Studies of Missense TREM2 Mutations in Human Stem Cell-
472 Derived Microglia. *Stem Cell Reports* **10**, 1294–1307 (2018).
473 76. Washer, S. J. *et al.* Single-cell transcriptomics defines an improved, validated monoculture
474 protocol for differentiation of human iPSC to microglia. *Sci. Rep.* **12**, 19454 (2022).
475 77. Nott, A. *et al.* Brain cell type-specific enhancer-promoter interactome maps and disease - risk
476 association. *Science* **366**, (2019).
477 78. Bégay, V., Baumeier, C., Zimmermann, K., Heuser, A. & Leutz, A. The C/EBPβ LIP isoform
478 rescues loss of C/EBPβ function in the mouse. *Sci. Rep.* **8**, 1–10 (2018).
479 79. Ryan, K. J. *et al.* A human microglia-like cellular model for assessing the effects of
480 neurodegenerative disease gene variants. *Sci. Transl. Med.* **9**, (2017).
481 80. Takaoka, A. *et al.* Protein tyrosine kinase Pyk2 mediates the Jak-dependent activation of MAPK
482 and Stat1 in IFN-γ, but not IFN-α, signaling. *EMBO J.* **18**, 2480–2488 (1999).
483 81. El Khoury, J. B. *et al.* CD36 Mediates the Innate Host Response to β-Amyloid. *J. Exp. Med.* **197**,
484 1657–1666 (2003).
485 82. Lee, J. K., Schuchman, E. H., Jin, H. K. & Bae, J.-S. Soluble CCL5 Derived from Bone Marrow-
486 Derived Mesenchymal Stem Cells and Activated by Amyloid β Ameliorates Alzheimer's Disease in
487 Mice by Recruiting Bone Marrow-Induced Microglia Immune Responses. *Stem Cells* **30**, 1544–1555
488 (2012).
489 83. Zhang, K. *et al.* CXCL1 Contributes to β-Amyloid-Induced Transendothelial Migration of
490 Monocytes in Alzheimer's Disease. *PLoS One* **8**, (2013).
491 84. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science*
492 **351**, 84–88 (2016).
493 85. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform.
494 *Bioinformatics* **26**, (2010).
495 86. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis.
496 *Nucleic Acids Res.* **44**, W160–W165 (2016).
497 87. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
498 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2013).
499 88. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and
500 microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).

501 89. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
502 RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
503 90. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential
504 expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
505 91. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions
506 of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
507 92. Buenrostro, J., Wu, B., Chang, H. & Greenleaf, W. ATAC-seq: A Method for Assaying Chromatin
508 Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1 (2015).
509 93. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078
510 (2009).
511 94. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, 1–9 (2008).
512 95. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
513 features. *Bioinformatics* **26**, 841–842 (2010).
514 96. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak
515 annotation, comparison and visualization. *Bioinformatics* **31**, (2015).
516 97. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**,
517 357.
518 98. Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for
519 CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 1–11 (2019).
520 99. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* **16**,
521 e2005970 (2018).
522 100. Coetzee, S. G., Coetzee, G. A. & Hazelett, D. J. motifbreakR: an R/Bioconductor package for
523 predicting variant effects at transcription factor binding sites. *Bioinformatics* **31**, 3847 (2015).
524 101. Haenseler, W. *et al.* A Highly Efficient Human Pluripotent Stem Cell Microglia Model Displays
525 a Neuronal-Co-culture-Specific Expression Profile and Inflammatory Response. *Stem Cell Reports*
526 **8**, 1727–1742 (2017).
527 102. Vaughan-Jackson, A. *et al.* Differentiation of human induced pluripotent stem cells to
528 authentic macrophages using a defined, serum-free, open-source medium. *Stem Cell Reports* **16**,
529 3093 (2021).
530 103. Jurga, A. M., Paleczna, M. & Kuter, K. Z. Overview of General and Discriminating Markers of
531 Differential Microglia Phenotypes. *Front. Cell. Neurosci.* **14**, 544457 (2020).
532 104. Murray, P. J. *et al.* Macrophage activation and polarization: nomenclature and experimental
533 guidelines. *Immunity* **41**, (2014).

534 # Methods

535 **Human induced pluripotent stem cells (hiPSC) culture**

536 KOLF2_C1 (https://hpscreg.eu/cell-line/WTSIi018-B-1) is a male hiPSC line generated as part of the
537 HipSci project (Cambridgeshire 1 NRES REC Reference 09/H0304/77, HMDMC 14/013) that has
538 been subcloned from a single cell to ensure homogeneity of the population. KOLF2_C1 cells were
539 cultured in TeSR E8 medium (StemCell Technologies) on Synthemax (Corning) (final amount 2
540 µg/cm^2) and routinely passaged 1:10 every 5 days using Gentle Cell Dissociation Reagent (StemCell
541 Technologies).
542

543 **Engineering rs28834970 C/C allele in hiPSCs**

544 The C/C risk allele at variant rs28834970 was engineered in hiPSCs by nucleofection of
545 ribonucleoprotein (RNP) complex containing full-length chemically modified synthetic guide RNA
546 and eSpCas9, along with a ssODN repair template, as previously described[72].
547 Briefly, eSpCas9[84] was expressed and purified from Escherichia coli using a His-tag. KOLF2_C1 cells
548 were cultured ahead of nucleofection in TeSR E8 medium (StemCell Technologies) on Synthemax
549 (Corning) (final amount 5 µg/cm^2). eSpCas9 (5µl, 20 µg) was mixed with full-length chemically
550 modified guide RNA (AGTGGAATTGTACAACACTG) (Synthego) (5 µl, 225 pmol) at room
551 temperature for 20 min for RNP complexes to form, followed by addition of the ssODN repair
552 template
553 (CTGGCAAGACTAATCTACTTTCTATTTTTATGGATCTGCCTTTTCTGGTCATTCCATATAAGTGGAATTGCA
554 CAACACTGTGGCCTTTCGCGACGGCTGCTTTCACTTAGCACAATGTTTTGAAACTTCCTCCATGTTGT) (5 µl,
555 500 pmol) just before the nucleofection. Cells were detached using Accutase (StemCell
556 Technologies) and dissociated into a single cell suspension. 1 × 10^6 cells were resuspended in P3
557 buffer (Lonza), mixed with the RNP/template complex and nucleofected using Lonza 4D-
558 Nucleofector with program CA137. After nucleofection cells were plated onto a 10 cm dish coated
559 with Synthemax (5 µg/cm^2) with TeSR-E8 supplemented with CloneR (StemCell Technologies). Cells
560 were cultured to 80% confluency, detached using Accutase and dissociated into single cells for
561 subcloning. 5000 cells were plated in a 10 cm dish and expanded, and colonies were picked for
562 genotyping and freezing. Colonies were screened for introduction of the homozygous C allele by
563 high throughput sequencing of an amplicon spanning the rs28834970 variant using an Illumina
564 MiSeq instrument. Final cell lines were validated using Sanger sequencing. Clones used in
565 downstream assays include: three homozygous (C/C) clones (named C8, C11 and D3), one unedited
566 (T/T) picked clone (B11) and the parental KOLF2_C1 line.
567

568 **hiPSC differentiation to macrophages**

569 iPSCs were differentiated to macrophages as previously described[73,74]. Briefly, KOLF2_C1 cells were
570 cultured in Essential E8 medium (Gibco) on Vitronectin (Gibco, 100x) and split onto a feeder layer
571 of Mouse Embryonic Fibroblasts (MEFs) (Merck) using Gentle Cell Dissociation Reagent (StemCell
572 Technologies). hiPSCs were cultured on the feeder layer in hiPSC base medium, containing
573 Advanced DMEM F12 (Gibco), KnockOut Serum Replacement (Gibco), 2mM Glutamax (Gibco), 100
574 U/ml penicillin/streptomycin (Gibco), 0.055 mM 2-Mercaptoethanol (Sigma) and 4ng/ml of
575 Recombinant Human FGF (R&D Systems), until the colonies were sufficiently large. HiPSC colonies
576 were lifted using a solution of equal amounts of Collagenase type IV (ThermoFisher Scientific) and
577 Dispase type II (ThermoFisher Scientific) for 15 minutes at 37°C. Detached colonies were plated in
578 10 cm low adherent dishes in hiPSC base medium without FGF to form Embryoid Bodies (EBs). After
579 five days, EBs were harvested and plated on gelatin coated 10 cm dishes in EB medium, containing
580 X-Vivo 15 medium (Lonza), 2 mM Glutamax (Gibco), 100 U/ml penicillin/streptomycin (Gibco),
581 0.055mM β-mercaptoethanol (Sigma), 50 ng/ml rhM-CSF (R&D Systems) and 25 ng/ml rhIL-3 (R&D
582 Systems). Macrophage precursor cells could be routinely harvested from dishes containing EBs
583 every 6/7 days for up to 3/4 months. Precursors were differentiated for 7 days by plating 15,000
584 cells/cm^2 in macrophage media, containing RPMI 1640 medium (Gibco), 10% FBS (Thermo Fisher

585 Scientific), 2 mM Glutamax (Gibco), 100 U/ml penicillin/streptomycin (Gibco) and 100 ng/ml rhM-
586 CSF (R&D Systems).
587
588 **HiPSCs differentiation to microglia**
589 HiPSCs were differentiated to microglia as previously described[75]. Briefly, KOLF2_C1 cells were
590 cultured in Essential E8 medium (Gibco) on Synthemax (Corning) (final amount 5 µg/cm^2). Two days
591 after passaging, cells were detached using Accutase (StemCell Technologies) and dissociated into
592 a single cell suspension. 10,000 hiPSCs were plated into each well of a 96-well plate in EB medium,
593 containing Essential E8 medium (Gibco), 10 uM Y-27632 (Dihydrochloride)(StemCell technologies),
594 50 ng/ml rhBMP-4 (Perprotech), 20 ng/ml rhSCF (Perprotech) and 50 ng/ml rhVEGF (Perprotech).
595 The 96-well plate was span down at 300xg for 3 minutes and incubated for 4/5 days for EB
596 formation. EBs were then transferred to tissue culture dishes coated in 0.1% gelatin (Sigma) and
597 cultured in Microglia Precursors media, containing X-Vivo 15 medium (Lonza), 2 mM Glutamax
598 (Gibco), 100 U/ml penicillin/streptomycin (Gibco), 0.055 mM β-mercaptoethanol (Sigma), 100
599 ng/ml rhM-CSF (Perprotech) and 25 ng/ml rhIL-3 (Cell Guidance Systems). Microglia precursor cells
600 could be routinely harvested from dishes containing EBs every 6/7 days for up to 3 months.
601 Precursors were differentiated for 7 days by plating at a density of 180,000 cells/cm^2 in microglia
602 media, containing RPMI 1640 medium (Gibco), 10% FBS (Thermo Fisher Scientific), 2 mM Glutamax
603 (Gibco), 100 U/ml penicillin/streptomycin (Gibco), 10 ng/ml rhGM-CSF (Perprotech) and 100 ng/ml
604 rhIL-34 (Perprotech).
605
606 **Fluorescence activated cell sorting analysis**
607 Microglia or macrophages were detached using Accutase (StemCell Technologies) and washed
608 twice in PBS. 1x10^5 cells were incubated with Human TruStain FcX (Fc Receptor Blocking Solution)
609 (Biolegend) for 30 minutes at 4°C in the dark. Microglia were stained with 5 ul each of PE-anti
610 human CD45 antibody clone HI30 (Biolegend) and APC-anti human CD11b antibody clone ICRF44
611 (Biolegend) for 30 minutes at 4°C. Macrophages were instead stained with 5 ul of either APC/Cy7-
612 anti human CD16 antibody clone 3G8 (Biolegend), APC-anti human CD206 antibody (BD
613 Biosciences) or PE-anti human CD14 antibody (BD Biosciences). Isotype controls were stained with
614 either 0.2 mg/ml APC-mouse IgG1 k(BD Biosciences), 0.2 mg/ml of PE-mouse IgG2A (BD
615 Biosciences) or 0.2 mg/ml APC/Cy7 mouse IgG1 k (BD Biosciences). Cells were then washed three
616 times in FACS buffer (5% FBS in D-PBS) and resuspended in 200 μl of DAPI solution (1 ug/ml in PBS)
617 and analysed. All samples were analysed using a BD LSRll instrument. An unpaired t-test was used
618 to test the difference between samples from cells with the T/T and the C/C allele at rs28834970.
619
620 **RNA extraction**
621 TRI Reagent (Zymo) was added direct on cell layers previously washed in PBS and mixed. RNA was
622 extracted using the Direct-zol RNA Miniprep kit (Zymo) following the manufacturer's protocol. The
623 optional in-column DNase digest was performed, and RNA was eluted in 50 µl DNAse/RNAse-free
624 water. We made cDNA from 1 µg RNA using Superscript IV (Thermo-Fisher) according to the
625 manufacturer's protocol.
626
627 **Reverse transcription**
628 1 µg RNA was reverse transcribed using Superscript IV (Thermo-Fisher) and random hexamers,
629 according to the manufacturer's protocol.
630
631 **RNA sequencing of hiPSC-derived macrophages and microglia**
632 Transcriptome libraries for hiPSC-derived macrophages and microglia were generated with the
633 Illumina TruSeq stranded RNAseq kit (polyA) and all samples were sequenced using Illumina HiSeq
634 with 50 million mapped reads on average for each sample.

635 Reads were mapped to the human genome (GRCh38/hg38) using the BWA software[85] and coverage
636 tracks were visualised on the UCSC genome browser using bigwig files, made with bamCoverage[86]
637 and normalised by Counts per Million (CPM). The number of reads per transcript were counted
638 using the featureCounts software[87] and a custom R script was used to calculate Transcripts per
639 Million (TPM). For RNAseq analysis of macrophages, TPMs were corrected for batch effects using
640 limma function limma::removeBatchEffect[88]. For microglia, all samples were differentiated and
641 sequenced in the same batch. Raw read counts per transcript were used to analyse differential
642 gene expression between cells with the T/T and the C/C allele at rs28834970, using the package
643 DESeq2[89]. DESeq2 was also used to generate PCA plots of variance stabilizing transformation (vst)
644 transformed counts. To investigate whether the rs28834970 variant had an effect on PTK2B
645 splicing, we used EdgeR[90] to calculate differential exon usage between microglia harbouring the
646 C/C and the T/T allele. Gene ontology (GO) analysis of differential expressed genes was carried out
647 using g:Profiler package in R[91].
648
649 **ATACseq of hiPSC-derived macrophages and microglia**
650 hiPSC-derived macrophages and microglia were detached using a solution of 4 mg/ml Lidocaine
651 hydrochloride monohydrate (Sigma) and 5mM EDTA (ThermoFisher scientific) for 15 minutes at
652 37°C. ATACseq was performed as previously described[92]. Briefly cells were lysed using sucrose
653 buffer containing 10 mM Tris-Cl pH 7.5, 3 mM $CaCl_2$, 2 mM $MgCl_2$, 0.32 M sucrose and
654 permeabilised using 0.5% Triton-X-100 (Sigma). Nuclei were recovered by spinning at 450xg for 5
655 minutes at 4°C and tagmentation was performed using Illumina Tagment DNA TDE1 Enzyme and
656 Buffer Kits for 30 minutes at 37°C37C. 5 μl of TDE1 enzyme was used per 100,000 cells. MinElute
657 PCR Purification Kit buffer PB (Qiagen) was added to stop the reaction and tagmented DNA was
658 purified and eluted in 10 μl of EB. The whole volume was PCR amplified and dual indexed Illumina
659 adapters were added (primer sequences from[92]). Amplified libraries were gel purified and
660 sequenced using an Illumina Novaseq 6000 with 30 million reads on average per sample.
661 Reads were mapped to the human genome (GRCh38/hg38) using the BWA software[85] and
662 duplicate reads were removed using picard tools MarkDuplicates. Coverage tracks were visualised
663 on the UCSC genome browser using bigwig files, made with bamCoverage[86] and normalised by
664 Counts per Million (CPM). The number of mapped reads per sample was determined and all files
665 were subsampled to the number of reads of the lowest sample, using samtools view[93]. ATACseq
666 peaks were called using MACS2[94] with flags --shift -100 --extsize 200 for window extension. A
667 merged file of all called peaks in all samples was created and the number of reads under each peak
668 was counted in all samples using bedtools coverageBed function[95]. Genome-wide analysis of
669 differential chromatin accessibility between cells harbouring the C/C and the T/T allele at
670 rs28834970 was performed using DESeq2[89]. DESeq2 was also used to generate PCA plots of
671 variance stabilizing transformation (vst) transformed counts. Peaks were annotated to genes using
672 CHIPseeker functiomannotatePeak[96].
673
674 **CUT & RUN assay**
675 Microglia were harvested using a solution of 4mg/ml Lidocaine hydrochloride monohydrate
676 (Sigma) and 5mM EDTA (ThermoFisher Scientific) for 15 minutes at 37°C37C. CUT&RUN for CEBPβ
677 was performed using CUTANA CUT&RUN Kit (EpiCypher) according to the manufacturer's
678 instructions. 400,000 microglia were used per sample and permeabilization was optimised using
679 Digitonin (EpiCypher) 0.01%. 1ug of C/EBP beta Antibody (H-7) sc-7962 (Santa Cruz Biotechnology)
680 was used per sample and one control sample was treated in parallel with 1 μl of CUTANA Rabbit
681 IgG CUT&RUN Negative Control Antibody. 0.5ng of E. coli Spike-in DNA (EpiCypher) was added to
682 each sample to normalise reads to control for experimental variability and sequencing depth.
683 Around 5 ng of purified CUT&RUN-enriched DNA was used to prepare NGS libraries using the
684 NEBNext Ultra II DNA Library Prep Kit (NEB) with NEBNext Multiplex Oligos for Illumina (Dual Index

685 Primers Set 1) (NEB). All libraries were sequenced 150bp paired end using an Illumina Novaseq
686 6000 with 15 million reads on average per sample.
687 Reads were trimmed to exclude Illumina adapter sequences using trim_galore and mapped to the
688 human genome (GRCh38/hg38) using the bowtie2 software[97]. The number of human mapped
689 reads per sample was determined and all files were subsampled to the number of reads of the
690 lowest sample, using samtools view[93]. Trimmed reads were also mapped to the E. coli genome
691 using bowtie2 and reads mapped to the human genome were normalised to spiked-in E.coli DNA
692 using a previously published script (https://github.com/Henikoff/Cut-and-
693 Run/blob/master/spike_in_calibration.csh) integrated into a custom bash script
694 (https://github.com/ericabello/PTK2B_rs28834970/blob/main/Cut_Run_CEBPb_micro/cmds_Cut
695 Run_CEBPb_micro.sh). Coverage tracks were visualised on the UCSC genome browser using bigwig
696 files made from E.coli normalised bedgraph files of fragments 1-1000bp. Bigwigs were made using
697 UCSC bedGraphToBigWig. Peaks in all samples containing the CEBPβ antibody were called using
698 the SEACR software[98] and the IgG negative control sample was used as input to identify the
699 threshold value at which the percentage of target versus IgG signal is maximized. A merged file of
700 all CEBPβ peaks in all samples was created and the number of reads under each peak was counted
701 in all samples using bedtools coverageBed function[95]. Genome-wide analysis of differential CEBPβ
702 binding between microglia harbouring the C/C and the T/T allele at rs28834970 was performed
703 using DESeq2[89].
704
705 **Digital droplet qPCR**
706 50 ng of cDNA per sample was mixed with 10 µl of ddPCR Supermix for probes (no dUTP) (Biorad),
707 1 µl of FAM-tagged TaqMan assay for detection of the gene of interest and 1 µl of VIC-tagged
708 Taqman assay for a housekeeping gene control. The housekeeping control gene was chosen to
709 match the level of expression of the gene of interest in each cell type.
710
711 Taqman assays list (ThermoFisher Scientific), containing probes and primers:
712 *Macrophages*

gene	FAM probe ID	control gene	VIC control probe ID
CCL1	Hs00171072_m1	TBP	Hs00427620_m1
CCL8	Hs00271615_m1	SDHA	Hs00188166_m1
CCL13	Hs00234646_m1	GAPDH	Hs99999905_m1
CCL5	Hs99999048_m1	TBP	Hs00427620_m1
CCL2	Hs00234140_m1	UBC	Hs00824723_m1
TRIM35	Hs01093979_m1	TBP	Hs00427620_m1
CLU	Hs00156548_m1	TBP	Hs00427620_m1
SCARA3	Hs00939871_m1	TBP	Hs00427620_m1
PTK2B	Hs01559703_m1	SDHA	Hs00188166_m1
HLA-DQA1	Hs03007426_mH	TBP	Hs00427620_m1
HLA-F	Hs01587840_m1	TBP	Hs00427620_m1

713
714 *Microglia*

gene	FAM probe ID	control gene	VIC control probe ID
CCL1	Hs00171072_m1	HPRT1	Hs02800695_m1
CCL22	Hs00171080_m1	TBP	Hs00427620_m1
CCL8	Hs00271615_m1	TBP	Hs00427620_m1
CXCL6	Hs00237017_m1	TBP	Hs00427620_m1
CCL13	Hs00234646_m1	UBC	Hs00824723_m1
CXCL2	Hs00601975_m1	TBP	Hs00427620_m1
IL18	Hs01038788_m1	TBP	Hs00427620_m1
CXCL3	Hs00171061_m1	TBP	Hs00427620_m1
CXCL1	Hs00236937_m1	SDHA	Hs00188166_m1
CXCL5	Hs00171085_m1	SDHA	Hs00188166_m1
CCL5	Hs99999048_m1	TBP	Hs00427620_m1
CCL2	Hs00234140_m1	GAPDH	Hs99999905_m1
TRIM35	Hs01093979_m1	TBP	Hs00427620_m1
CLU	Hs00156548_m1	TBP	Hs00427620_m1
SCARA3	Hs00939871_m1	TBP	Hs00427620_m1
PTK2B	Hs00169444_m1	TBP	Hs00427620_m1
HLA-DQA1	Hs03007426_mH	TBP	Hs00427620_m1
HLA-F	Hs01587840_m1	TBP	Hs00427620_m1

715
716 Droplets were generated for each reaction using a QX200 Droplet Generator (Biorad) and
717 transferred to 96 well plates. Sealed plates were incubated in a thermocycler and cDNA was PCR
718 amplified with the following steps: 95°C for 10 min then 95°C for 30 sec and 60°C for 1 min for 40
719 cycles and finally 98°C for 10 min. The resulting FAM and VIC fluorescence was read using a QX200
720 Droplet Reader (Biorad) and the fraction of positive droplets in the sample was determined.
721 Poisson statistics were used to determine the absolute number of copies per µl in each sample for
722 both target and housekeeping control genes. The number of copies of target gene per sample was
723 normalised by the control gene for each reaction and an unpaired t-test was used to test the
724 difference between samples from cells with the T/T and the C/C allele at rs28834970.
725
726

727 **Microglia stimulation and Luminex assay**
728 hiPSC-derived microglia at day 6 of differentiation were stimulated with IFNγ (R&D Systems) 200
729 ng/ml or Lipopolysaccharide (LPS) (Invitrogen) 100 ng/ml for 24 hours. Culture plates were spun at
730 4,000 rpm for 20 min at 4°C and media collected. Harvested media was diluted 1:20 for
731 unstimulated controls and 1:250 for stimulated samples. Luminex assay was performed using
732 ProcartaPlex Human Basic Kit (Invitrogen) and a panel of ProcartaPlex Human Simplex (Invitrogen)
733 beads, according to the manufacturer's recommendations. Plates were read using a Luminex
734 MAGPIX instrument. Media from cells with the T/T and C/C allele were ran on the same plate,
735 allowing direct comparison of each protein concentration across genotypes. An unpaired t-test was
736 used to test the difference in concentration (pg/ml) between different samples. The propagation
737 of error for log fold change values of chemokine concentration between unstimulated and
738 stimulated microglia was calculated using the Python package uncertainties
739 (https://pythonhosted.org/uncertainties/).
740
741 **Microglia stimulation and migration assay**
742 HiPSCs-derived microglia at day 7 of differentiation were stimulated with IFNγ (R&D Systems) 100
743 ng/ml for 72 hours. At day 10, cells were detached using a solution of 4 mg/ml Lidocaine
744 hydrochloride monohydrate (Sigma) and 5 mM EDTA (Gibco) and 5.5×10^4 cells in 100 μl
745 macrophage media were seeded onto each transwell (PET with 5 µm pores, Sarstedt) in a 24-well
746 plate. Cells were incubated for 15 minutes to settle and 600 μl of macrophage media containing 3
747 nM human recombinant C5a was added beneath the transwells. The plate was incubated for 10
748 hours to allow cell migration. The transwell was then rinsed in PBS, transferred to a fresh 24-well
749 plate, and fixed with 4% paraformaldehyde for 20 min at RT. Cells on either side of the transwell
750 membrane were stained with NucBlue (Invitrogen) and imaged with an EVOS FL Auto automated
751 microscope (Thermo Fisher) set up to take images of each transwell in full (top image). The
752 transwells were swabbed with a cotton wool bud to remove cells on the top surface, leaving behind
753 only migrated cells, transferred into a fresh plate and imaged again with the same settings (bottom
754 image). Cell counting was performed with CellProfiler 3.0 software[99] and the percentage of
755 migrated cells per transwell was calculated as: (no. cells in bottom image) ÷ (no. cells in top
756 image) × 100. Stimulations were performed in triplicate and the average for each sample across
757 wells was taken. An unpaired t-test was used to test the difference between different samples.
758

759 ## Data availability

760 Raw sequencing data is available at ENA https://www.ebi.ac.uk/ena/browser/search with
761 accession number ERP024123 for ATACseq and Cut&Run data and ERP024128 for RNAseq
762 data. A description of the sample IDs can be found in Supplementary Table 4.
763 All analysis scripts are available at https://github.com/ericabello/PTK2B_rs28834970.
764

765 ## Acknowledgements

766 This work was funded by OpenTargets [OTAR037] and Wellcome [206194]. We thank the
767 Sanger Institute flow cytometry core facility for flow cytometry analysis. Addgene
768 plasmids, etc. We acknowledge Scientific Operations at the Sanger Institute for support
769 in next generation sequencing and quality control. We thank members of the Bassett lab
770 for helpful discussions and support.
771
772
773

774 ## Author contributions

775 A.B. and E.B. conceived the work and designed experiments, K.L. made the hiPSC isogenic
776 cell lines, E.B., S.I., J.S. executed the experiments, E.B., K.A., N.K., J.S., N.I.P. analysed
777 data, E.B. and A.B. wrote the manuscript with input from all other authors.
778

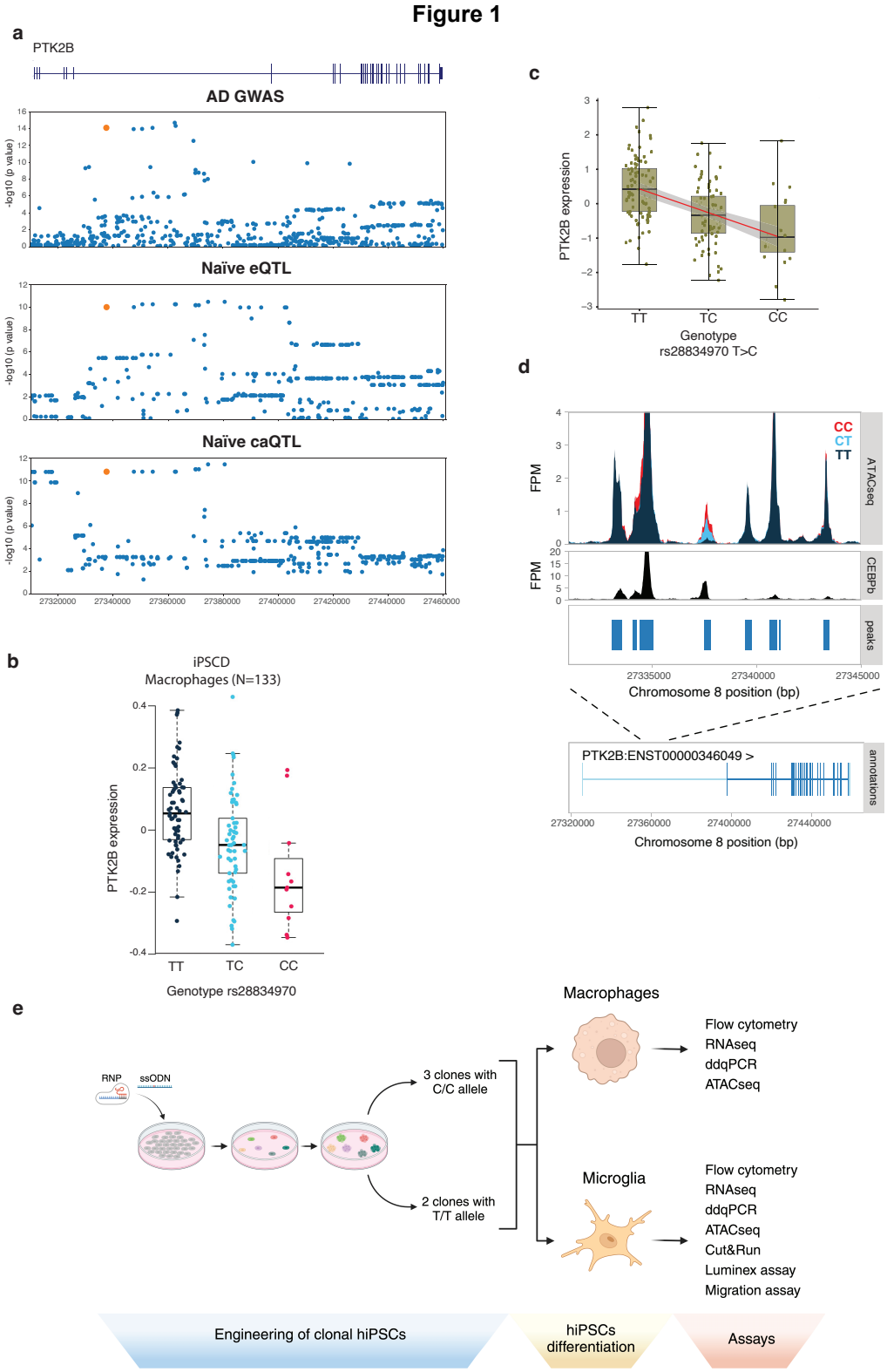
779 ## Competing interests

780 A.B. is a founder of EnsoCell therapeutics. N.I.P. was an employee of GSK at the time the
781 manuscript was submitted. J.S. was an employee of Illumina at the time the manuscript
782 was submitted.
783

784 ## Materials and Correspondance

785 All correspondence and requests for materials should be addressed to Andrew Bassett
786 (ab42@sanger.ac.uk) or Erica Bello (eb956@cam.ac.uk).
787
788

789 **Figures and legends**



790
791 **Figure 1 - The rs28834970 variant in PTK2B is predicted to modulate an intronic enhancer**
792 A) *PTK2B* gene structure and Manhattan plots for Alzheimer's Disease (AD) GWAS hits (top
793 panel), eQTL in naïve iPSC-derived macrophages (second panel) and caQTL in naïve iPSC-
794 derived macrophages (bottom panel). P-values replotted from[69].

B) Boxplots showing the expression of the *PTK2B* gene stratified by the rs28834970 genotype in iPSC-derived macrophages. Data replotted from[25]. The y-axis shows normalised expression levels (log TPM value) and each dot shows the expression level of a single sample.

C) Boxplots showing the expression of the *PTK2B* gene stratified by the rs28834970 genotype in naïve iPSC-derived macrophages. Data replotted from[70]. The y-axis shows normalised expression levels (log TPM value) and each dot shows the expression level of a single sample.

D) ATAC-seq fragment coverage in iPSC-derived macrophages stratified by the rs28834970 genotype (top panel) and CEBPβ ChIP-seq fragment coverage and peaks in primary human macrophages (middle panel). Data replotted from[69]. The bottom panel shows the structure of the *PTK2B* gene.

E) Overview of the experimental design. Created with BioRender.com.

Figure 2

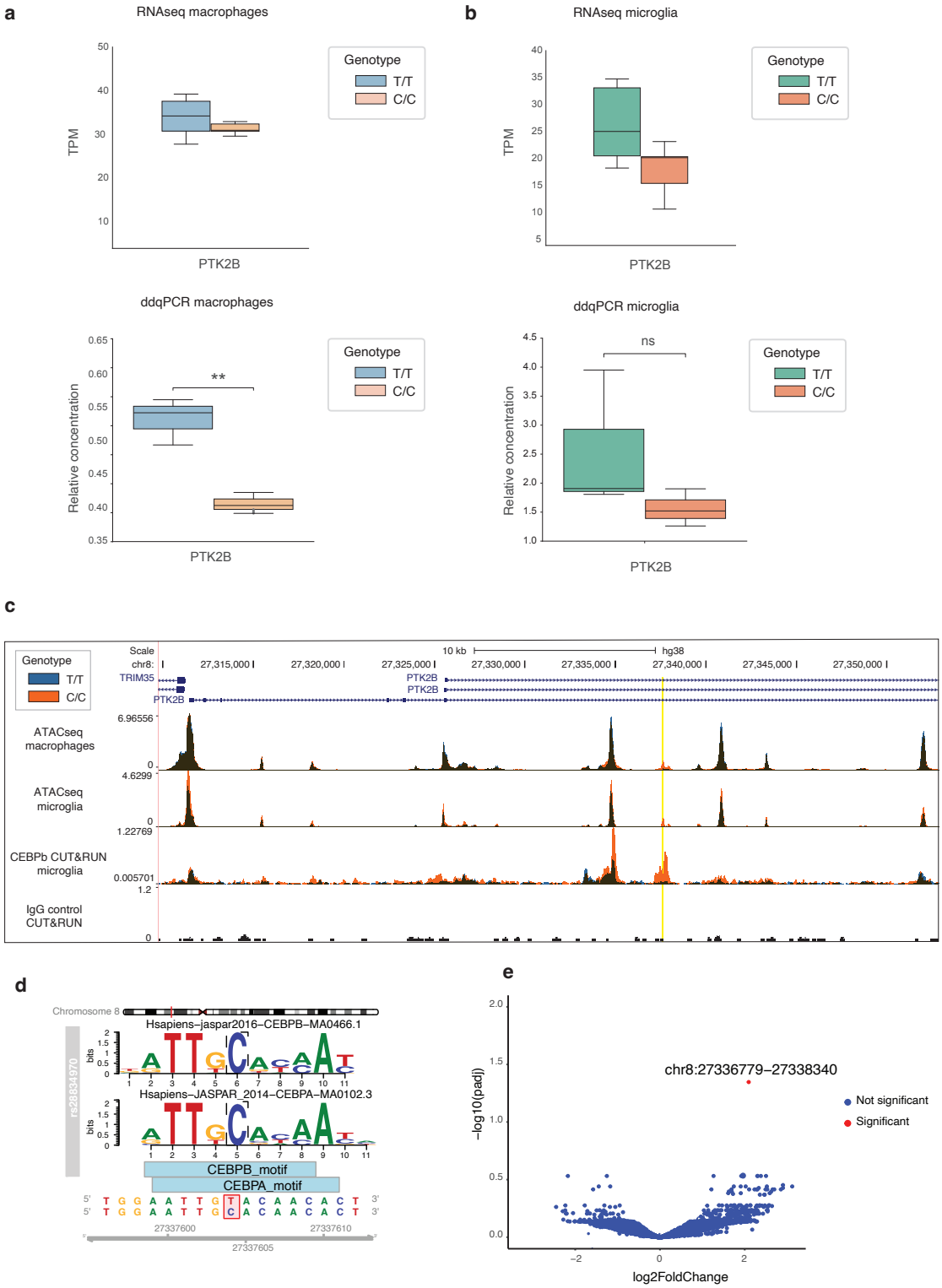


Figure 2 - Isogenic edited cell lines show that rs28834970 effects *PTK2B* expression through effects on chromatin accessibility and CEBPβ binding at the intronic enhancer in iPSC-derived macrophages and microglia.

A) Boxplots showing *PTK2B* expression in macrophages harbouring the T/T or C/C allele at rs28834970 measured by RNA-seq (top) and droplet digital q-PCR (ddqPCR) (bottom). Top plot: TPM= transcripts per million. Bottom plot: Relative concentration = copies/ul relative to housekeeping gene (SDHA), n=3 and **p<0.01, one-way paired ANOVA.

B) Boxplots showing *PTK2B* expression in microglia harbouring the T/T or C/C allele at rs28834970 measured by RNA-seq (top) and droplet digital q-PCR (ddqPCR) (bottom). Top plot: TPM= transcripts per million. Bottom plot: Relative concentration = copies/ul relative to housekeeping gene (TBP), n=3 and **p<0.01, one-way paired ANOVA. ns= non-significant

C) Coverage tracks of cells with T/T (orange) and C/C (blue) allele at rs28834970. The tracks are: ATAC-seq in macrophages (upper track), ATAC-seq in microglia (second track), CUT&RUN for CEBPβ in microglia (third track) and IgG negative control for CUT&RUN (bottom track). The rs28834970 variant in *PTK2B* intron is highlighted in yellow.

D) Effect of rs28834970 SNP on transcription factor binding probability. Position probability matrices are shown for CEBPβ (top) and CEBPα (middle) as well as the position of the binding motifs on the *PTK2B* sequence (bottom). The plot was generated using motifbreakerR[100].

E) Volcano plot showing fold changes of CEBPβ binding between microglia harbouring the C/C and the T/T allele for all detectable peaks, measured by CUT&RUN. Blue dots indicate peaks with no significant change while red dots represent significantly increased peaks (logFC>1 and p<0.05). The only significantly increased CEBPβ peak found spans the rs28834970 variant in the *PTK2B* intron.

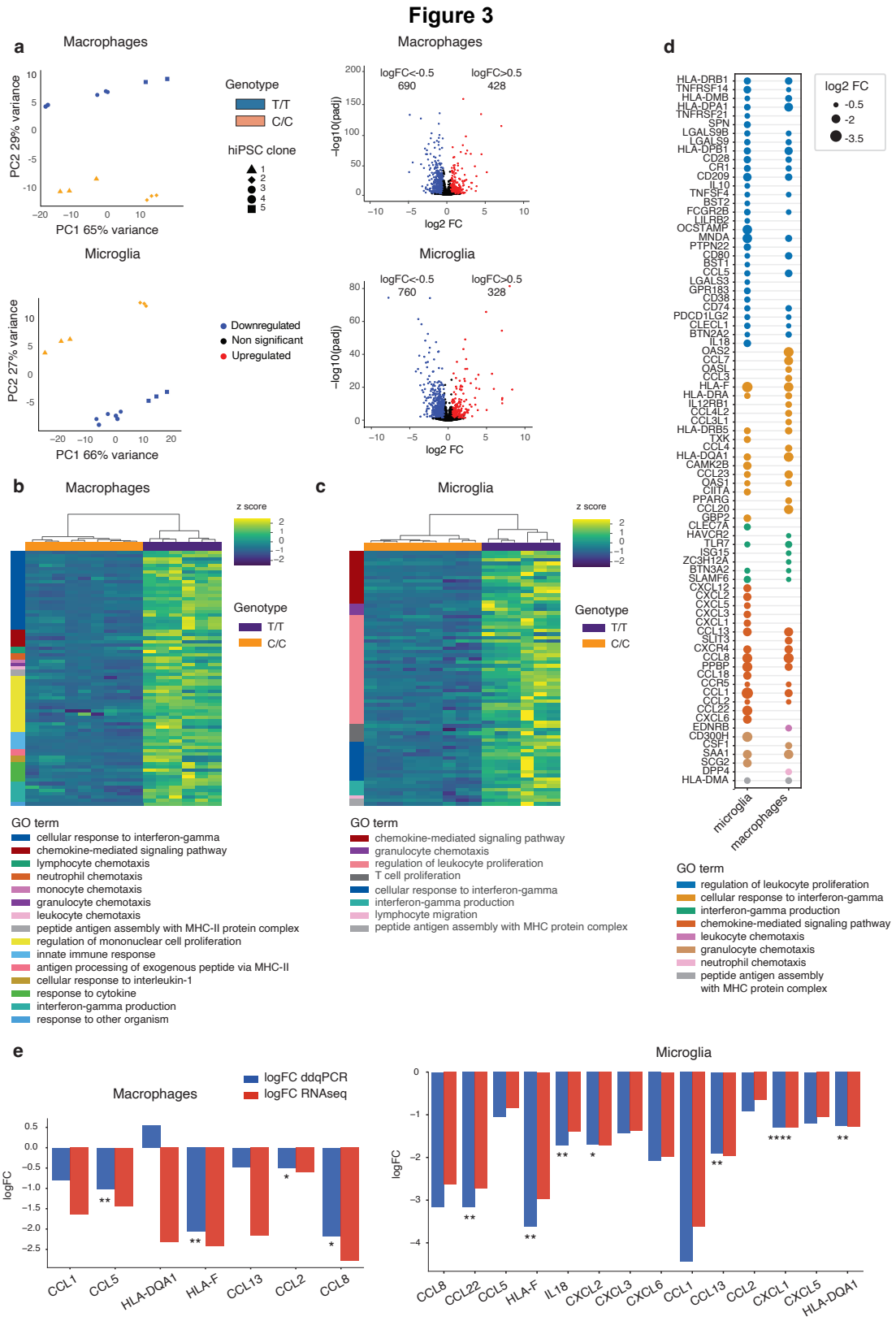


Figure 3 - The rs28834970 variant affects global gene expression in iPSC-derived macrophages and microglia

838 A) Results of differential expression analysis of macrophages (upper panels) and microglia
839 (lower panels) harbouring the C/C allele versus the corresponding iPSC-derived cells
840 harbouring the T/T allele at rs28834970. Left panels show Principal component analysis
841 (PCA) of transcriptomes of cells differentiated from two hiPSC clones with the T/T allele
842 and three clones with the C/C allele. Right panels are a volcano plot showing the log2 fold-
843 change (FC) between cells with the C/C and the T/T allele and corresponding p values for
844 all transcribed genes. Significantly upregulated genes (logFC>0.5 and p<0.05) are shown as
845 red dots, significantly downregulated genes (logFC<-0.5 and p<0.05) are shown as blue
846 dots and non-significant genes as black dots.
847 B) Heatmaps showing relative expression (z-scores) of DE genes found in gene ontology
848 (GO) analysis in macrophages with the T/T (orange) and the C/C (purple) allele at
849 rs28834970. GO pathways identified are also annotated.
850 C) Heatmaps showing relative expression (z-scores) of DE genes found in gene ontology
851 (GO) analysis in microglia with the T/T (orange) and the C/C (purple) allele at rs28834970.
852 GO pathways identified are also annotated.
853 D) Dot plot showing logFC of downregulated genes belonging to the top eight GO pathways
854 in macrophages and microglia.
855 E) Bar graph showing logFC of downregulated chemokines and HLA genes in macrophages
856 (left) and microglia (right) measured by RNAseq and ddqPCR. ddqPCR: n=3 *p<0.05,
857 **p<0.01, ***p<0.001, ****p<0.0001 by one-way paired ANOVA.
858

Figure 4

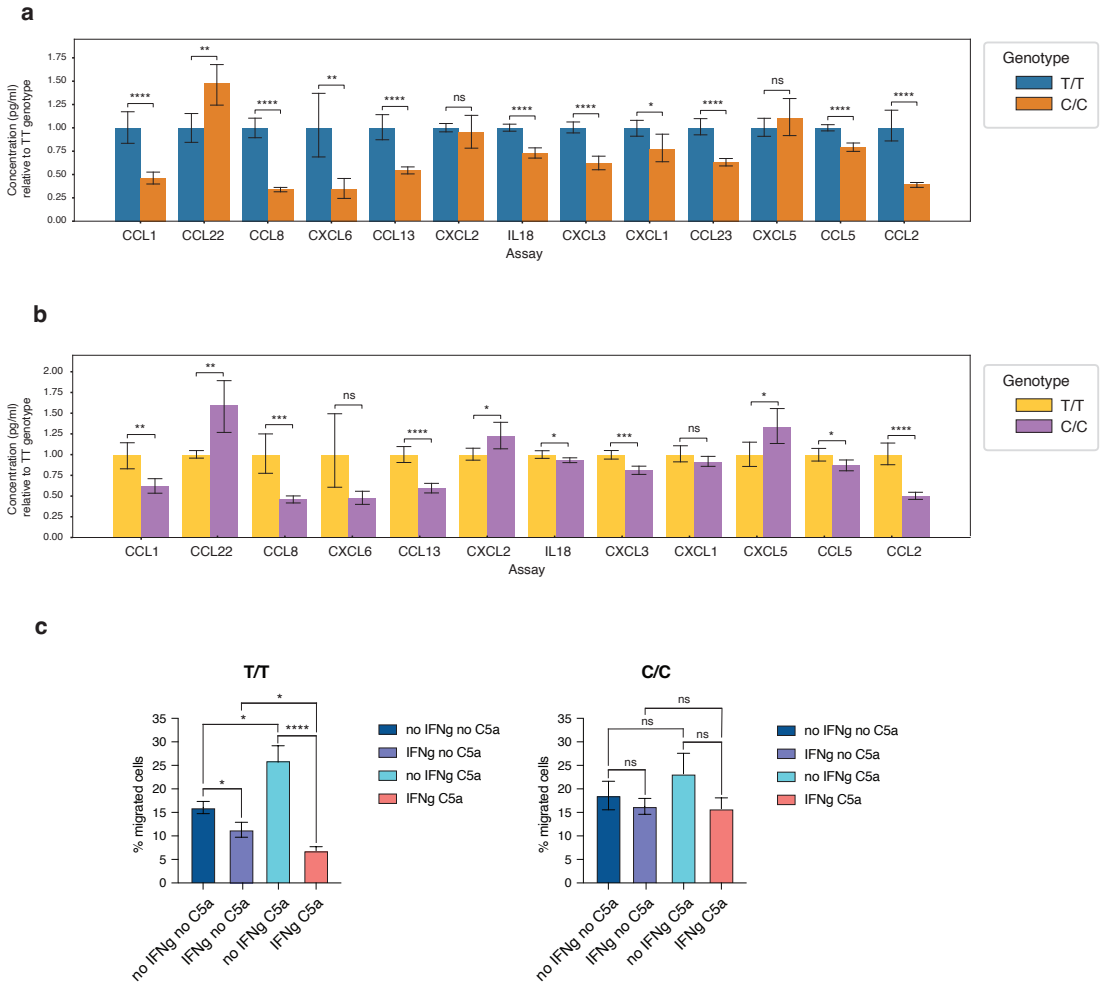


Figure 4 - The rs28834970 variant causes changes in chemokine production and migration in iPSC-derived microglia

A) Concentration of downregulated chemokines in unstimulated microglia with the C/C allele at rs28834970 relative to the concentration in cells with the T/T allele, measured by Luminex assay. Results are shown as the mean ± SEM of independent biological replicates (n=3). ns= non-significant, * p<0.05, **p<0.01, ***p<0.001, ****p<0.0001, unpaired t-test.

B) Concentration of downregulated chemokines in microglia with the C/C allele at rs28834970 relative to the concentration in cells with the T/T allele, stimulated with IFNγ. Results are shown as the mean ± SEM of independent biological replicates (n=3). ns= non-significant, * p<0.05, **p<0.01, ***p<0.001, ****p<0.0001, unpaired t-test.

C) Bar graph showing the percentage of migrated microglia harbouring the T/T allele (left) or the C/C allele (right) either unstimulated or stimulated with IFNγ, and with or without the chemoattractant C5a. Results are shown as the mean ± SEM of independent biological replicates (n=3). ns= non-significant, * p<0.05, **p<0.01, ***p<0.001, unpaired t-test.