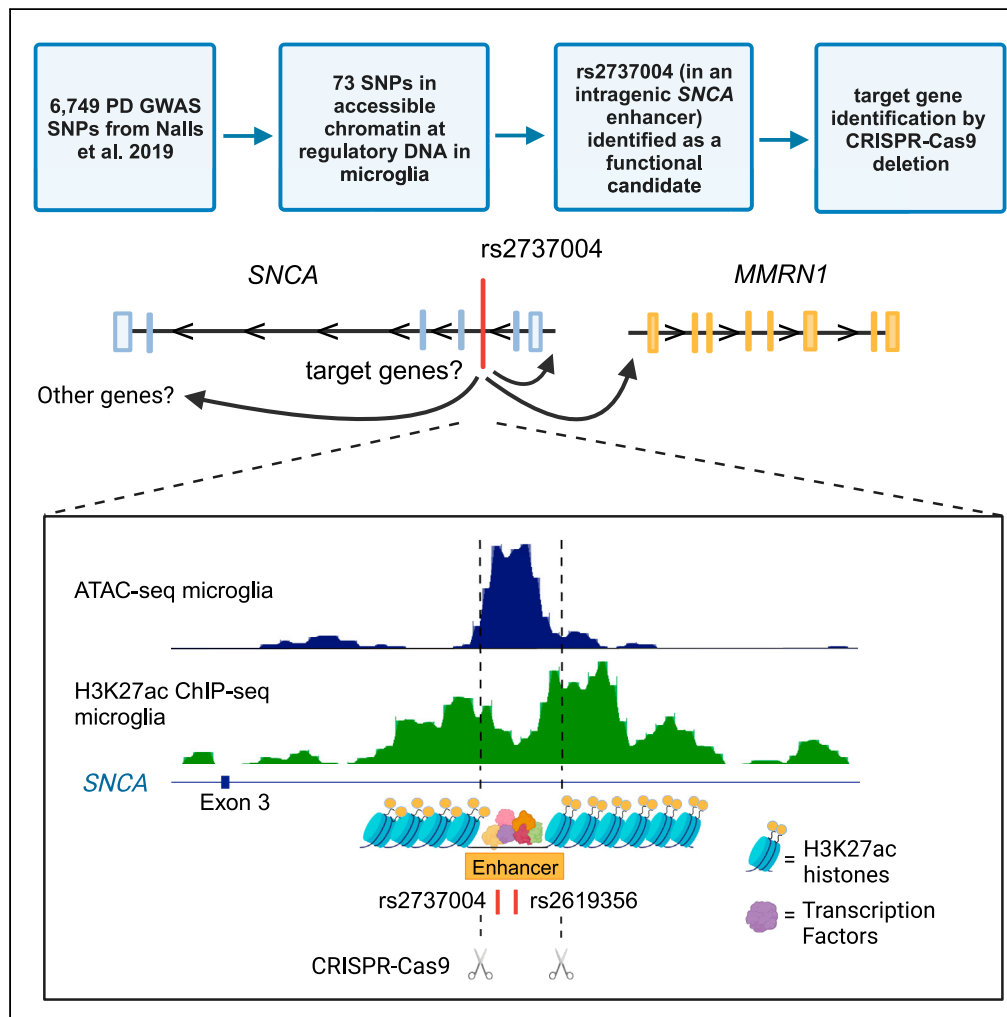


Article

# Parkinson's disease risk enhancers in microglia



Alix Booms, Steven E. Pierce, Edwin J.C. van der Schans, Gerhard A. Coetzee

[alix.booms@vai.org](mailto:alix.booms@vai.org)

**Highlights**

73 functional candidate risk loci were identified in microglia

Two PD risk SNPs reside in an intragenic microglia enhancer in SNCA

Deletion of the enhancer led to decreased SNCA and MMRN1 expression

Additional potential target genes were up-/downregulated by the enhancer deletion

Booms et al., iScience 27, 108921  
February 16, 2024 © 2024 The Authors.  
<https://doi.org/10.1016/j.isci.2024.108921>



## Article

## Parkinson's disease risk enhancers in microglia

Alix Booms,<sup>1,2,3,\*</sup> Steven E. Pierce,<sup>1</sup> Edwin J.C. van der Schans,<sup>1</sup> and Gerhard A. Coetsee<sup>1</sup>

## SUMMARY

Genome-wide association studies have identified thousands of single nucleotide polymorphisms that associate with increased risk for Parkinson's disease (PD), but the functions of most of them are unknown. Using assay for transposase-accessible chromatin (ATAC) and H3K27ac chromatin immunoprecipitation (ChIP) sequencing data, we identified 73 regulatory elements in microglia that overlap PD risk SNPs. To determine the target genes of a "risk enhancer" within intron two of *SNCA*, we used CRISPR-Cas9 to delete the open chromatin region where two PD risk SNPs reside. The loss of the enhancer led to reduced expression of multiple genes including *SNCA* and the adjacent gene *MMRN1*. It also led to expression changes of genes involved in glucose metabolism, a process that is known to be altered in PD patients. Our work expands the role of *SNCA* in PD and provides a connection between PD-associated genetic variants and underlying biology that points to a risk mechanism in microglia.

## INTRODUCTION

Genetic studies suggest that a significant portion of Parkinson's disease (PD) risk is heritable (estimated to be up to 36%).<sup>1</sup> Unlike some rare disorders caused by highly penetrant mutations in a small number of genes, PD genome-wide association studies (GWASs) have uncovered thousands of low-penetrance single-nucleotide polymorphisms (SNPs) with more modest influences on disease risk. It is hypothesized that their small effects play a role in PD predisposition through subtle changes in gene expression over the course of an entire lifespan. There are thus ongoing efforts to understand the effects of these (and other) DNA risk variants prior to disease onset.

The findings of GWASs have been challenging to interpret as not all PD-associated SNPs are biologically functional in relevant cellular and developmental contexts. Moreover, most of them are in non-coding DNA.<sup>2</sup> Some variants within enhancers and promoters are known to influence gene expression regulation.<sup>3</sup> However, the genes whose expression they target are not immediately apparent.<sup>4</sup> Thus, the first challenges in the field are to 1) determine which SNPs are imposing PD risk by altering biology and 2) identify mechanisms by which each allele of a risk SNP leads to biological differences.

A large portion of PD risk SNPs are enriched in regulatory enhancers and promoters across multiple cell types, implying that PD susceptibility may in part be attributed to genetic variation that impacts the regulation of cell-type-specific genes and cellular processes.<sup>5</sup> Microglia are the resident macrophages in the brain and are known contributors to neuroinflammation in PD.<sup>6</sup> PD-associated variants reside in regulatory elements or are in and around PD-associated genes in these cells.<sup>7,8</sup> However, the variants that are functional and how they influence microglia processes are largely unknown. It has been demonstrated that SNPs within regulatory elements alter transcription factor binding and gene expression.<sup>3,9,10</sup> The effects of subtle genotype-dependent gene expression may impact microglia functions and ultimately disease risk over the course of a lifetime.

The goal of our study was to identify PD-associated SNPs that potentially function via allele-dependent regulation of gene expression in microglia. We first mapped open chromatin in induced pluripotent stem cell (iPSC)-derived microglia using assay for transposase accessible chromatin with sequencing (ATAC-seq). We combined our data with published ATAC-seq data from primary microglia<sup>11</sup> to create a consensus list of open chromatin regions. We then overlapped these consensus regions with published H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) datasets from primary *ex vivo* microglia tissue<sup>11</sup> to demarcate active enhancers and promoters. These regions overlapped 73 out of 6,749 "SNPs of interest" published in the latest GWAS meta-analysis.<sup>1</sup> We report these as candidates for in-depth mechanistic evaluation in microglia.<sup>12</sup>

For thorough functional analysis, we chose to focus on one of our top candidate regulatory elements, an intragenic "risk enhancer" at *SNCA*, defined by its overlap with a PD risk SNP. GWASs of PD have uncovered many risk variants at the *SNCA* locus which make up at least three independent association signals.<sup>13,14</sup> However, distinct functional variants at this locus have yet to be identified in microglia. In this study we report on two variants, rs2737004 and rs2619356, that we believe to be functional in microglia. They are in linkage disequilibrium (LD) with a PD-association signal spanning the 5' end of *SNCA* into *MMRN1*. CRISPR/Cas9 deletion of the open chromatin region, containing these variants, led to reduced expression of *SNCA* and the adjacent gene *MMRN1*, confirming a regulatory effect on nearby genes. In addition, there was a small subset of differentially expressed genes involved in cell-cycle regulation and glucose metabolism, which are two linked processes

<sup>1</sup>Center for Neurodegenerative Science, Van Andel Institute, Grand Rapids, MI 49503, USA

<sup>2</sup>Van Andel Institute graduate student, Grand Rapids, MI 49503, USA

<sup>3</sup>Lead contact

\*Correspondence: [alix.booms@vai.org](mailto:alix.booms@vai.org)  
<https://doi.org/10.1016/j.isci.2024.108921>



involved in microglia inflammatory responses. Our work underscores the importance of evaluating genetic risk on a cell-type-specific basis and implicates *SNCA*, and a specific association signal there, as an important risk locus in microglia.

## RESULTS

### Regions of accessible chromatin are consistent between iPSC-derived and primary microglia

We evaluated genetic risk in microglia using the iPSC-derived microglia model developed by McQuade et al.<sup>15</sup> Following the differentiation of iPSCs to mature microglia, we performed ATAC-seq to map genome-wide regions of accessible chromatin ( $n = 4$ ). iPSC-derived model systems represent a promising alternative to primary tissue due to their relative ease of creation and experimental manipulation as well as the ability to follow their trajectory through differentiation. However, culturing cells in an *in vitro* environment may affect the chromatin landscape.<sup>11</sup> Therefore, to augment our newly generated ATAC-seq data from iPSC-derived microglia, we reanalyzed ATAC-seq data from 13 *ex vivo* primary microglia samples published by the Glass laboratory.<sup>11</sup> For these samples, microglia were isolated from the brain tissue of 13 different patients, both male and female, ranging in age from 2 to 17 years. The samples were derived from various regions of the brain, including the temporal cortex, the frontal cortex, the occipital cortex, and the cerebellum. We combined the datasets to identify active chromatin regions in iPSC-derived microglia shared with a heterogeneous population of primary microglia. Presumably, these common loci are relevant to a broad range of microglia types.

Many ATAC peaks are shared between iPSC-derived microglia and primary microglia. Out of 73,276 peaks present in iPSC-derived microglia, 60,139 (~82%) overlap with primary microglia (Figure S1A). An example of peak consistency at a particular locus is shown in Figure S1B around *CX3CR1*, a gene known to be expressed in microglia. For subsequent data analyses, we used regions of open chromatin that were present in both iPSC-derived and primary microglia.

### 73 out of 6,749 PD risk SNPs overlap functional regulatory DNA in microglia

SNPs that reside in active regulatory DNA are known to function by disrupting transcription factor binding, leading to changes in gene expression.<sup>16</sup> To identify such SNPs, we first overlaid the location of PD risk "SNPs of interest" from Nalls et al.<sup>17</sup> with consensus iPSC-derived and primary microglia ATAC-seq peaks. We also combined that data with published H3K27ac ChIP-seq data from 3 primary *ex vivo* microglia samples<sup>11</sup> to narrow down ATAC-seq peaks within regulatory DNA. Out of 6,749 SNPs, 73 were in ATAC-seq peaks and were in or within 100 bp of an H3K27ac ChIP-seq peak (Table S1).

From the 73 SNPs in ATAC peaks, we highlight six top candidate risk SNPs (Table 1, and related Table S1) ranked by GWAS p value. Other than rs4889599, we observed that these SNPs are in ATAC peaks that are flanked by H3K27ac ChIP-seq peaks (Figure S2). This pattern is what we typically consider optimal when evaluating candidate loci because it displays nucleosome displacement surrounded by acetylated histone marks, indicative of transcriptional activity and transcription factor occupancy. Indeed, a MotifbreakR analysis showed that 51 out of the 73 SNPs in ATAC-seq peaks are predicted to have an allele-specific preference for one or more transcription factors expressed in microglia (Table S2).<sup>18</sup>

### A "risk haplotype" resides in an intragenic *SNCA* enhancer

*SNCA* is a well-known PD risk gene that has multiple PD GWAS association signals at its locus, yet there are virtually no studies that have dissected genetic risk in microglia at this region. As an in-depth follow-up we focused on an enhancer in intron two of *SNCA* that overlaps one of our top-ranking candidate risk SNP, rs2737004. Our initial goal was to zoom in on this region to determine if there are other SNPs that are in LD with rs2737004. NCBI's LDproxy tool shows one additional SNP (rs2619356) that is in LD with rs2737004 and maps to the open chromatin region of the same *SNCA* enhancer (Figure 1B). The LDpair tool showed that the two SNPs are in LD ( $D' = 1$ ), where the G allele of rs2737004 always co-segregates with the T allele of rs2619356 (Figure 1A). Herein we refer to this allele pair as the "risk haplotype". Note that the  $R^2$  value is low because the allele frequencies in each haplotype are not equal (i.e., one allele is less common than the other). Our previous analysis showed that SNPs with unequal allele frequencies tend to have larger effects.<sup>12</sup> In this case the minor allele frequency of rs2737004 is not proportionate to the major allele (G = 10.7% and A = 89.3% in all NCBI populations). Furthermore, the G allele of rs2737004 is only present with the T allele of rs2619356 in a smaller subset of all populations (10.74%). We therefore predict that the risk haplotype could have relatively large effects, diluted when large populations are analyzed. It is important to note that rs2737004 is not near the most significant GWAS-associated signal tagged by rs356182 (p value =  $1.85 \times 10^{82}$ ). The two SNPs are also not strongly correlated ( $D' = 0.6786$ ,  $R^2 = 0.0816$ ). Instead, rs2737004 appears to be in LD with a different independent signal tagged by rs763443 ( $D' = 0.9484$ ) that is located at the opposite end of the gene. These risk signals are reviewed elsewhere by Pihlstrom et al.<sup>14</sup>

### The *SNCA* risk enhancer shows evidence of functionality via 3D chromatin interactions and transcription factor binding

Judging by proximity, *SNCA* is the most likely target gene of the intragenic enhancer encompassing rs2737004 and rs2619356. However, enhancers are known to control multiple genes in *cis* and *trans*.<sup>19</sup> To profile close-range interactions of the risk enhancer, we used published PLAC-seq (proximity ligation-assisted ChIP-seq) data from human microglia, neurons, and oligodendrocytes<sup>20</sup> (Figure 2A). Compared to neurons and oligodendrocytes, microglia have a unique interaction profile characterized by frequent contact with *MMRN1* and *GPRIN3* (denoted by red ovals in Figure 2A). In microglia, most of the contact points correspond with the presence of an H3K27ac signal, indicating that these 3D

**Table 1. Top candidate risk SNPs that overlap active regulatory DNA**

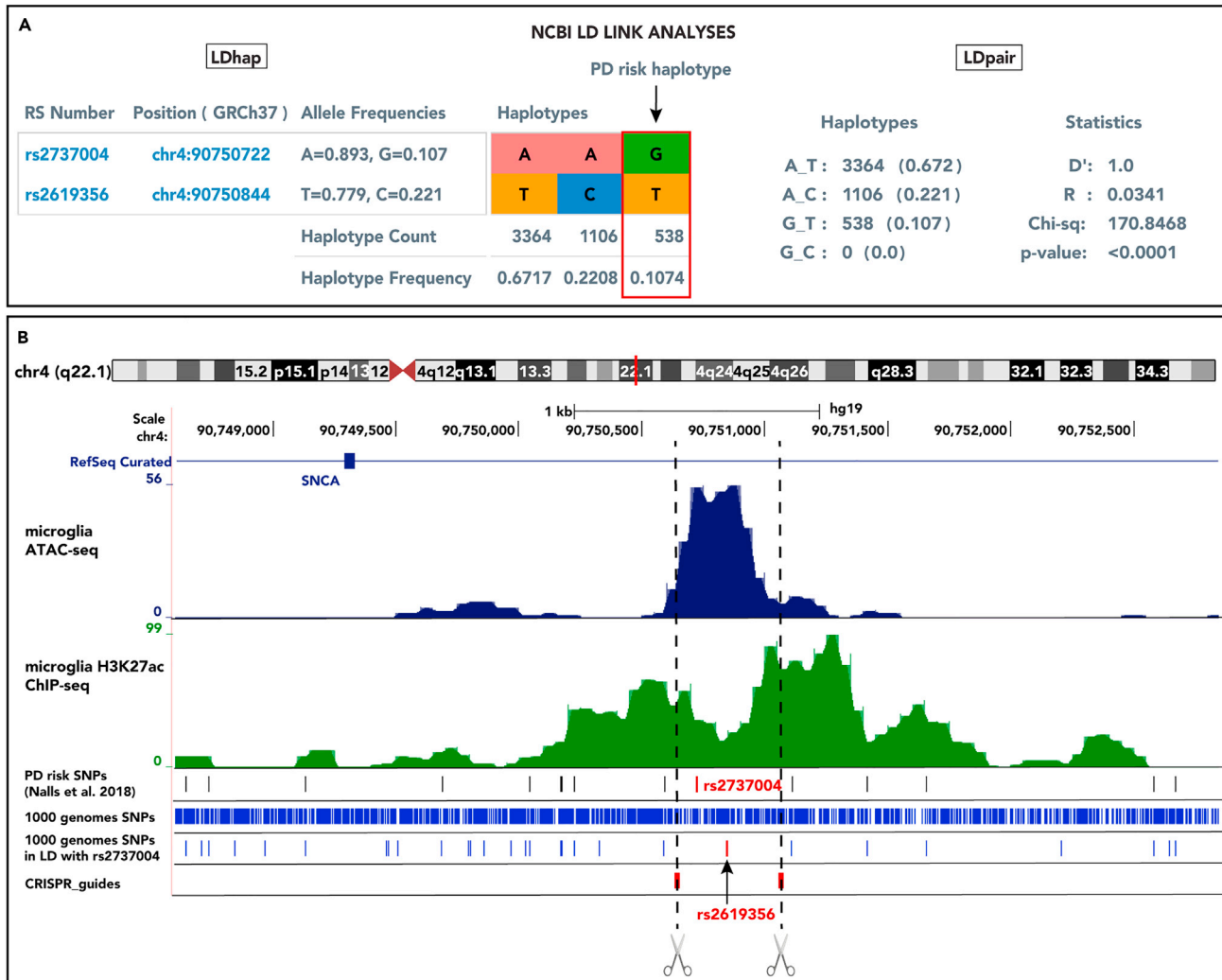
SNP	Closest gene	GWAS p value	Function of gene	Role in neurodegeneration
rs12726330	SLC50A1 (SWEET1)	2.208x10 <sup>-27</sup>	glucose transport across the cell membrane <sup>45</sup>	No studies have directly evaluated SLC50A1, but decreased expression of glucose transporters correlates with hypometabolism of glucose in neurodegenerative diseases. <sup>46</sup>
rs2737004	SNCA	7.60x10 <sup>-11</sup>	regulation of dopamine release, induces fibrilization of microtubule-associated protein tau, and protection against apoptosis by inhibiting p53 and caspase-3 activation <sup>47</sup>	It is mostly known to form toxic aggregates in neurons, but, in iPSC-derived macrophages from patients with an SNCA triplication, phagocytosis and cytokine release was impaired. <sup>48,49</sup>
rs144814361	BAG3	9.07x10 <sup>-11</sup>	co-chaperone that interacts with Hsp70 to prevent apoptosis during aging and under conditions of stress <sup>49</sup>	The protein is protective against neurodegeneration by preventing NLRP3 activation and inflammation in microglia. <sup>50</sup>
rs3813020	FBXL19	2.05x10 <sup>-10</sup>	regulates the ubiquitination and degradation of inflammatory cytokines, <sup>51</sup> regulates RhoA signaling <sup>52</sup>	Its role in neurodegeneration has not directly been tested. However, loss of RhoA is associated with microglia dysfunction and neurodegeneration. <sup>53</sup>
rs4889599	SETD1A	7.34x10 <sup>-10</sup>	histone methyltransferase that regulates transcriptional programming during embryogenesis <sup>54</sup>	Loss of function is associated with neurodevelopment disorders and dysfunctional neuronal metabolism. <sup>55</sup> It is a GWAS-identified PD risk gene, but its functional involvement in the disease is unknown. <sup>56</sup>
rs823114 rs7536483	NUCKS1	4.35x10 <sup>-9</sup>	involved in cell growth, proliferation, DNA repair, metabolism, and inflammatory and immune responses <sup>57</sup>	The gene is part of the PARK16 locus (one of the first to be identified by PD GWAS). Its expression is downregulated in PD patients, but its causal mechanisms are unknown. <sup>58</sup>

Table of top-ranking SNPs (based on GWAS p value) in open chromatin at active regulatory regions of DNA in microglia. See also [Table S1](#) and [Figure S2](#).

interactions happen between the portion of *SNCA* that contains the risk enhancer and active regulatory DNA at *MMRN1* and *GPRIN3*. The interaction profile is different in neurons, where some of the contact points are present within the H3K27ac signal. However, the interactions appear to be more localized around *SNCA* and the intergenic region between *SNCA* and *GPRIN1*. In oligodendrocytes there are few 3D interactions, all of which are between *SNCA* and *GPRIN1* (data not shown). This suggests that the *SNCA* risk enhancer functions differently in microglia than in neurons or oligodendrocytes and highlights the cell-specific activity of the risk enhancer.

Both SNPs of the "risk haplotype" could have a synergistic influence on gene expression levels of *SNCA* or other genes via allele-specific binding to transcription factors. As a first approximation for such a mechanism, we used MotifbreakR<sup>18</sup> to find transcription factors that are predicted to have differences in binding strength to the protective and risk alleles of rs2737004 and rs2619356. The most frequent motif altered by rs2737004 genotype was CTCF, which shows a preference for G (the risk allele) ([Figure 2B](#); [Table S2](#)). RAD21, a core subunit of the cohesin complex, also shows a preference for the G allele ([Figure 2B](#); [Table S2](#)). To provide supporting evidence of CTCF and RAD21 binding at rs2737004, we searched ChIP-seq data from ReMap.<sup>21</sup> Although microglia are not a cell type in the ReMap database, multiple other cell types, including peripheral macrophages (a related cell type), show binding of CTCF at rs2737004 ([Figure 2C](#) orange bars). There is also evidence of RAD21 binding at the same location ([Figure 2C](#) blue bars). CTCF and RAD21 work together to mediate 3D chromatin structure, forming loops between enhancers and promoters to facilitate chromatin accessibility and gene expression.<sup>22,23</sup> A plausible mechanism leading to increased risk for PD could thus involve changes in 3D chromatin organization that impact the expression of multiple genes.

The MotifbreakR analyses also showed that rs2619356 influences the affinity for many transcription factors ([Table S2](#)). Of the transcription factors that have a strong preference for the T-allele (part of the risk haplotype), CEBP proteins, specifically CEBPB and CEBPE, appeared most frequently (3 motifs) ([Figure 2B](#); [Table S2](#)). There was no ReMap data for CEBPE. However, CEBPE binds to a motif at rs2619356 in multiple cell types, including macrophages. CEBPB is a basic-leucine zipper transcription factor that regulates pro-inflammatory responses in microglia.<sup>24</sup> It has also been shown to bind at *SNCA* and promote its expression in a neuroblastoma cell line.<sup>25</sup> Others have linked *SNCA* dysregulation to allele-specific transcription factor occupancy.<sup>3,26</sup> It is hypothesized that this mechanism leads to subtle increases in *SNCA* expression over a long period of time, contributing to alpha-synuclein aggregation later in life. These data point to a similar mechanism. However, this analysis is a preliminary step in demonstrating the functionality of rs2619356 and rs2737004. Moving forward we focused on identifying the target genes of the risk enhancer created by CRISPR/Cas9-mediated deletion of the open chromatin region of the risk enhancer.



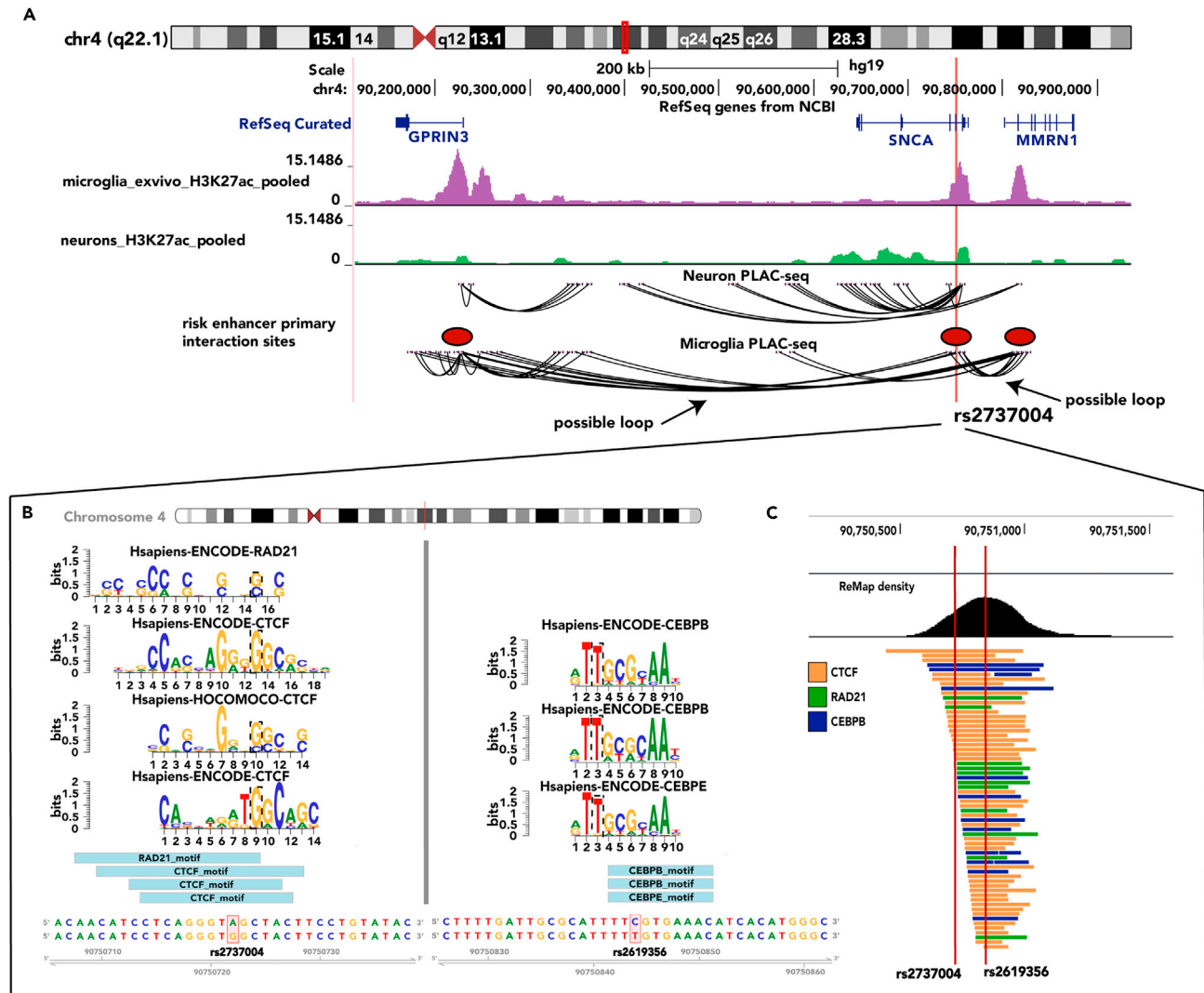
**Figure 1. A “risk haplotype” resides in an intragenic SNCA enhancer**

(A) Correlation analysis of the alleles of rs2737004 and rs2619356 using NCBI’s LDlink tools. The LDhap results show the frequency of each allele of the SNPs individually in all populations from NCBI. The colored boxes represent the haplotypes observed in the same population, with their counts and frequencies displayed below. The LDpair analysis reports the calculated statistics for linkage disequilibrium ( $D'$  and  $R^2$ ) and the “goodness-of-fit” (chi-square and p value), which indicates the degree that the observed haplotype frequencies deviate from the expected allele frequencies.

(B) Genome browser view of microglia ATAC- and H3K27ac ChIP-seq signals plotted with the locations of PD risk SNPs, all SNPs from 1,000 genomes, and the SNPs that are in LD with rs2737004. Rs2619356 was the only other SNP to overlap the ATAC-seq peak. Dotted lines represent the location of CRISPR/Cas9 guides designed to delete the 439 base pair region encompassing both SNPs in the “risk haplotype.”

### The SNCA risk enhancer controls expression of SNCA, MMRN1, and a network of additional genes in microglia

When defining risk genes, a common approach is to implicate the nearest genes to the SNP. Whereas this approach has provided substantial insight into PD-related pathways, it may not reveal the full extent of gene targets because an enhancer may affect genes multiple kilobases away on linear DNA or even on different chromosomes.<sup>27</sup> To determine the target genes of the enhancer in which the risk haplotype resides, we created a 439 bp deletion of the open chromatin region containing rs2737004 and rs2619356 (Figure 1B). We then performed bulk RNA sequencing (RNA-seq) at three time points across differentiation of microglia: day 0 iPSCs, day 12 hematopoietic progenitors (HPCs), and day 40 microglia. Sample names, their collection times, and total number of technical and biological replicates can be found in Figure S4B. There were no significant differences between standard microglia marker genes *AIF1* (*IBA1*), *TMEM119*, *CD11b*, and *P2RY12* in wild-type compared to edited cell lines (Figure S5C). Note that *TMEM119* mRNA expression was low in microglia. The use of this marker as a robust indication of microglia has been recently challenged due to its variability, both increased and decreased, depending on activation status.<sup>28</sup> We also confirmed protein expression by immunocytochemistry (ICC) of *TMEM119* and *IBA1* in all cell lines (Figure S5B). We did not do a formal quantification due to inconsistencies related to timing and cell losses during the staining procedure.



**Figure 2. The SNCA risk enhancer shows evidence of functionality via 3D chromatin interactions and transcription factor binding**

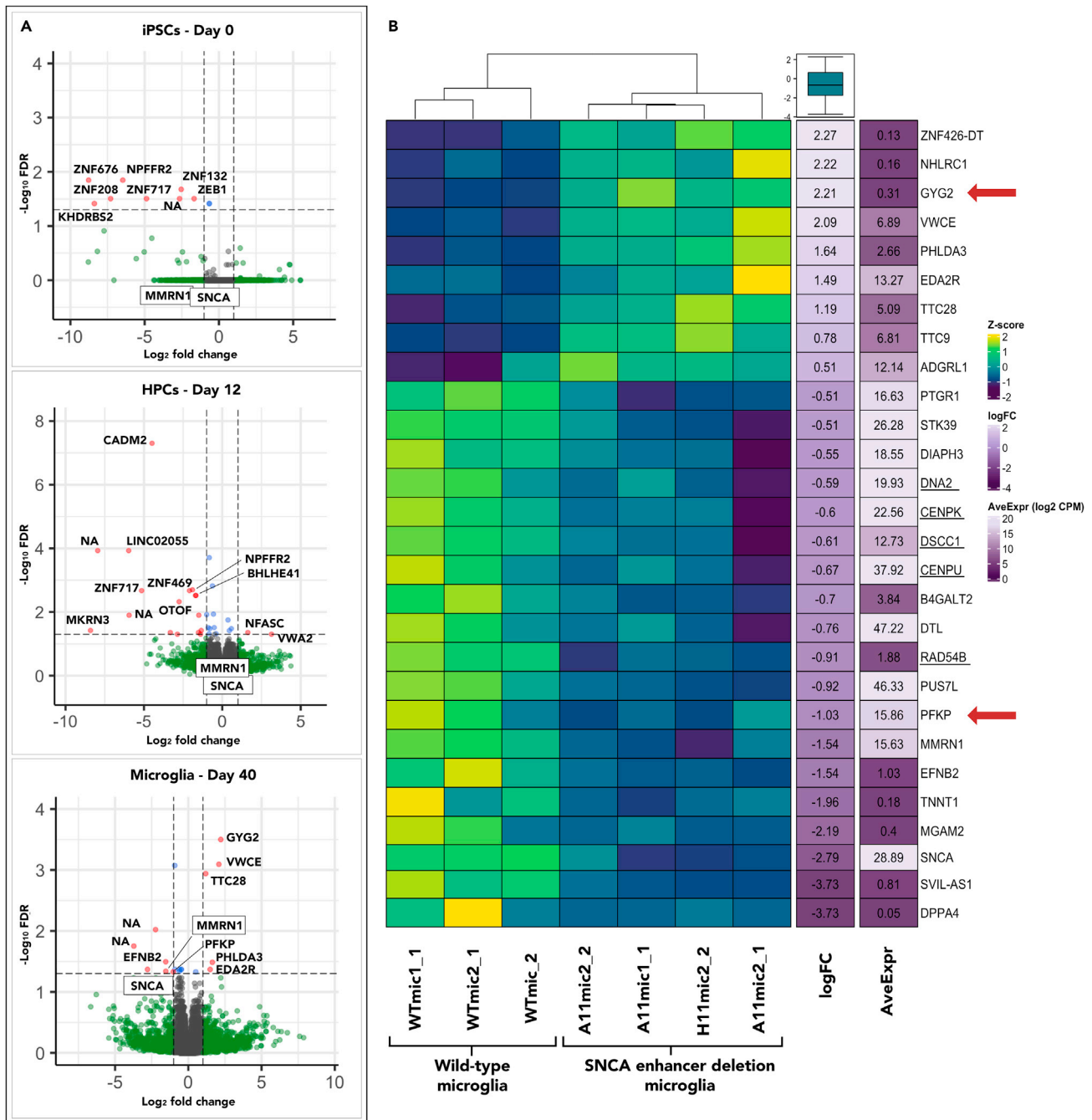
(A) H3K27ac ChIP-seq tracks are displayed for microglia (purple) and neurons (green). Below are tracks showing PLAC-seq data for the same cell types. The red ovals denote primary interaction sites, in microglia, of the risk enhancer where rs2727004 is located.

(B) MotifbreakR results showing transcription factor (TF) binding motifs of the TFs that have preference for the alleles of the risk haplotype (G for rs2737004 (left) and T for rs2619356 (right)). The letter size represents the results of the positional weight matrix that measures the frequency that the transcription factor binds to that nucleotide. In that same plot, the dashed black box demarcates the position of the SNP. The light blue boxes below represent the positions of the transcription factor binding motifs relative to the SNP's genomic position, demarcated with the red box.

(C) Remap ChIP-seq data for the transcription factors displayed in part B. The red lines show the position of each SNP within the ChIP-seq peak.

Looking across time points, we observed that the number of genes influenced by the enhancer deletion increased with the differentiation time course (Figure 3A and related Table S1). There was also little overlap between differentially expressed genes at each stage, suggesting that either the enhancer is not as active at earlier time points or it controls different sets of genes at each stage in differentiation. Interestingly, SNCA and MMRN1 expression significantly decreased in microglia but was unchanged in iPSCs and HPCs (Figure 3A), indicating a unique role for the risk enhancer in controlling expression of SNCA and MMRN1 in microglia. Our findings also corroborate the PLAC-seq data showing that MMRN1, in addition to SNCA, is a likely target of the risk enhancer in microglia. The PLAC-seq data indicated that GPRIN1 could be a target gene. However, its expression was not significantly different at any time points after the enhancer deletion.

Genes with a differential expression false discovery rate (FDR) of  $\leq 0.1$  in iPSCs and HPCs showed no enrichment of pathways identified by GONet gene ontology (GO) analysis tool.<sup>29</sup> Using the same cutoff in microglia, the only pathway to show enrichment was DNA conformation change (FDR = 1.2e-2), which is described by the Gene Ontology Resource as "a cellular process that results in a change in the spatial configuration of a DNA molecule." A conformational change can bend DNA or alter the twist, writhe, or linking number of a DNA molecule.<sup>30</sup> Seven



**Figure 3. The SNCA risk enhancer controls the expression of SNCA, MMRN1, and a network of additional genes in microglia**

(A) EdgeR glmFit results comparing wild-type to SNCA-deletion cell lines across three time points in differentiation. Dotted lines represent a Log<sub>2</sub> fold change of 2 and a false discover rate (FDR (Benjamini-Hochberg)) of 0.05. Gene name labels were manually added for consistency and clarity. See also Table S3 for results. (B). Heatmap of Z scores calculated from TMM-normalized log<sub>2</sub> counts per million (log<sub>2</sub> CPM). The genes in the heatmap represent those that have a Benjamini-Hochberg FDR cutoff of less than 0.1 (less stringent than used for the volcano plots in part A), and a fold change (FC) of over 1.4 (log FC > 0.5). The final plot was made by ranking the genes by log FC and taking the 15 genes at the top and bottom of that list.

out of the 42 genes, *CENPK*, *DNA2*, *HELLS*, *DSCC1*, *CENPU*, *RAD54B*, and *MCM8*, were a part of this network. The genes that had an FDR value  $\leq 0.05$  are displayed and underlined on the heatmap in Figure 3B. Looking more specifically, they are known to be involved in DNA replication and are likely indicative of changes in the cell cycle. This was confirmed using GO enrichment from the Gene Ontology Resource (FDR = 2.66E-03).<sup>30</sup> We also observed an inverse relationship between GYG2 (Glycogenin 2) and PFKP (Phosphofructokinase) expression

(Figure 3B red arrows). *GYG2*, a gene involved in glycogen biosynthesis, was most significantly upregulated while *PFKP*, a gene involved with the conversion of fructose 6-phosphate to fructose 1,6-bisphosphate at the beginning stages of glycolysis, was downregulated. This, along with changes in cell-cycle genes, suggests that loss of the enhancer could be affecting glucose metabolism and possibly shifting glucose utilization to glycogen storage rather than glycolysis. This is a phenotype that we plan to explore in future experiments.

## DISCUSSION

Here we aimed to determine a subset of PD-associated risk SNPs located in regions of active regulatory DNA in microglia and to identify functional risk SNPs in this cell type. In doing so, we substantially narrowed down the 6,749 PD-associated "SNPs of interest" from the latest PD meta-analysis to a more tractable list of 73. We chose one candidate risk locus, *SNCA*, for more in-depth evaluation based on its overlap with a top-ranking candidate SNP, rs2737004 (GWAS  $p$  value =  $7.6 \times 10^{-11}$ , OR = 1.14), in addition to the relevance of *SNCA* to PD risk, which is not well understood in microglia.

Multiple PD GWASs have reported independent association signals at the *SNCA* locus, but it is still unclear which variants are functional around these signals.<sup>1,14,31</sup> The top-ranking GWAS hit, rs356182 ( $p$  value =  $1.85 \times 10^{-82}$ ), is located at the 3' end of *SNCA* in a regulatory element identified in neurons.<sup>26</sup> Our lab previously demonstrated that heterozygous deletion of rs356182 causes changes in the expression of *SNCA* and thousands of additional genes, many of which are related to neuronal differentiation.<sup>26</sup> However, that same 3' enhancer is not active in microglia and none of the risk variants near rs356182 overlap microglia H3K27ac and ATAC signals. Alternatively, our analysis points to a potential functional SNP in microglia, rs2737004, that appears to segregate independently of rs356182 ( $D' = 0.6786$ ,  $R^2 = 0.0816$ ). Rs2737004 is linked to a separate independent GWAS signal, near the 5' end of *SNCA*.<sup>13,14</sup> It is in LD with the top tag SNP rs763443 ( $D' = 0.9484$ ). We speculate that the risk signal tagged by rs763443 represents a group of functional variants in microglia (or other immune cells), and our analysis pinpoints rs2737004 as a top functional candidate. In contrast, the risk signal at rs356182 may represent functional variants in other cells such as neurons. In the larger context of PD, we imagine a scenario where multiple nearby genetic risk signals each function through *SNCA*, but these separate signals represent biology relevant to different cell types. Thus, each set of regulated genes is unique, resulting in dysregulation of different cellular processes. For these reasons, we believe that, even though rs356182 carries the strongest association with PD risk, it is not the most relevant candidate for follow-up in microglia and more focus should be placed on understanding the function of PD-associated variation at the independent signal at the 5' end of *SNCA*.

Upon further evaluation of the *SNCA* locus, we located an additional SNP, rs2619356, in the risk enhancer that is in strong LD ( $D' = 1$ ) with rs2737004. Interestingly, one version of the haplotype is non-existent in the study population and the risk allele of rs2737004 only appears with the T allele of rs2619356, indicating co-segregation/linkage of the two alleles. Rs2619356 was not published in the original list of "SNPs of interest" likely due to the use of  $R^2$  as a measure of LD, which results in the exclusion of SNPs with allele frequencies that are not close to 50%. In terms of our approach to include functional SNPs based on their location in open chromatin at active regulatory DNA, rs2619356 is positioned in the center of the open chromatin region and transcription factor binding ReMap signal (Figure 2C). It is also predicted to have allele-specific preference for multiple transcription factors (Table S2), making it an equally relevant candidate for mechanistic follow-up.

A plausible mechanism of the risk enhancer could involve alterations in chromatin looping mediated by allele-specific CTCF and RAD21 binding. The G allele of rs2737004 has a stronger preference for CTCF and RAD21, which could promote the stability of the loop to facilitate *SNCA* and *MMRN1* expression. In 3 out of the 4 tissues profiled (breast, esophagus, and pituitary) by the Genotype-Tissue Expression project (GTEx), the GG genotype of rs2737004 is associated with increased expression of *SNCA*. The presence of CEBP within the motif for rs2619356 and its preference for the allele of the risk haplotype supports this idea, as CEBP has also been shown to promote the expression of *SNCA*. Studies, primarily in neurons, demonstrate that elevated expression of *SNCA* leads to alpha-synuclein aggregation and cellular dysfunction. However, the function of endogenous *SNCA* in microglia is not well known. *SNCA* has multiple roles that could require time-specific or varying levels of *SNCA* expression in different cell types. For example, alpha-synuclein is highly expressed in neurons and required for many important neuronal functions such as differentiation and synaptic transmission.<sup>32</sup> Alpha-synuclein has more recently been shown to facilitate immune responses and thus may be lowly expressed under homeostatic conditions in immune cells like microglia but becomes upregulated on a temporal basis in response to stress or infection.<sup>33,34</sup> In different cellular contexts, *SNCA* may therefore be regulated differently via cell-type-specific enhancers. This could explain the differences we observed in topology and regulatory signals like H3K27ac at *SNCA* when comparing microglia to neurons and provides a hypothetical scenario for how PD-associated variants within regulatory elements increase disease risk in distinct cell populations. However, this has not been directly demonstrated and how *SNCA* regulation is carried out in different cell types and different contexts such as viral infection and stress is still an early area of research.

Enhancer looping to target promoters is one of the critical aspects of proper gene regulation. Although the resolution of PLAC-seq data is not precise enough to distinguish whether the interaction of *MMRN1* is with the risk enhancer, the promoter of *SNCA*, or both, there does appear to be evidence for loop formation between *SNCA* and *MMRN1* (Figure 2A). This points to *MMRN1* as a potential target gene. We further believe that *MMRN1* is a plausible target, as deletion of the enhancer led to loss of *MMRN1* expression in addition to reduced *SNCA* expression. Whether the expression changes in the remaining set of genes are due to alterations in primary enhancer interactions or secondary downstream effects is still an open question.

*SNCA* expression changes, both up and down, have been associated with PD risk, but fewer studies have focused on how *MMRN1* relates to PD risk. *MMRN1* is mostly known to function in platelets to aid in coagulation, but there is limited understanding of its physiological functions.<sup>35</sup> Although the gene has been mostly studied in the context of cancer metastasis, a transcriptome-wide association analysis recently



identified *MMRN1* as a gene whose expression associates with PD risk.<sup>36</sup> *MMRN1* genetic abnormalities have also been found in autosomal dominant PD.<sup>37</sup> Surprisingly, *SNCA* and *MMRN1* expression was not significantly different in iPSCs or HPCs, whereas in microglia, deletion of the enhancer led to a loss in their expression. We hypothesize that the enhancer promotes or maintains *SNCA* and *MMRN1* expression as the cells differentiate into microglia. This suggests a biologically important role for both *SNCA* and *MMRN1* in microglia, but more studies are needed to corroborate this hypothesis and to understand their function in PD.

To evaluate whether our differentially expressed gene set contained any genes that are known to be dysregulated in PD patients, we looked for overlap with published gene sets from cells with genetic abnormalities in *SNCA*.<sup>38</sup> We confirmed overlap of 5 genes in our dataset with a gene set from A53T *SNCA* mutant dopaminergic neurons. In addition to *SNCA*, *MMRN1*, *DTL*, and *DPPA4* were upregulated whereas *PFKP* was downregulated in comparison to wild-type control cells. In a PD patient-derived dopaminergic cell line with an *SNCA* triplication, *SNCA* was upregulated, but *VWCE*, *EDA2R*, *PUS7L*, and *MGAM2* were downregulated. This suggests that changes in *SNCA* may affect a common set of genes in dopaminergic neurons and microglia. However, the changes we observed in our model may not be completely related to *SNCA* expression changes but more related to the loss of the enhancer that controls a specific set of genes, including *SNCA*, that are relevant to microglia function.

Two other genes stood out due to their involvement in glucose metabolism, which is a key process in microglia that controls their activation in response to inflammation.<sup>39</sup> The most upregulated gene, *GYG2*, is involved in glycogen biosynthesis. We also observed a loss of *PFKP*, which is a critical regulatory enzyme in glycolysis.<sup>40</sup> Our GO analysis identifying cell cycle as a process affected by 7 of the other differentially expressed genes supports a role for this metabolic phenotype in PD risk as metabolism and cell cycle are highly interconnected.<sup>41</sup> On a related note, out of the 73 SNPs in microglia regulatory DNA, the one with the most significant GWAS p value is at *SLC50A1* (Table 1; Figure S2), a glucose transporter whose function has never been studied in microglia. There have been clinical findings showing that PD patients have reduced glucose metabolism at early stages of the disease.<sup>42</sup> Strikingly, altered glucose metabolism in diabetes patients was also found to increase the chances for developing PD by 30%.<sup>43</sup> Our results and the findings of others justify the need for more studies to understand how dysfunctional glucose metabolism in microglia leads to increased risk for PD as this pathway may be a promising therapeutic target in PD.

To the best of our knowledge this is the first functional evaluation of a “risk enhancer” near the PD-association signal at the 5′ end of *SNCA*. How *SNCA*, *MMRN1*, and other genes are regulated by this enhancer may play an important part in PD pathogenesis by impacting inflammatory functions in microglia. Our data also provide a starting point for dissecting genetic risk at other loci and demonstrate the importance of careful evaluation of PD-associated variants on a cell-type-specific basis. We advocate for more post-GWAS testing of these risk variants to make sense of the genetic contribution to increased PD risk. There are currently no treatments to modify the progression of PD. Additional studies that build on our findings will help understand the complex genetic etiology of PD and identify alternative disease-modifying targets.

### Limitations of the study

Although we demonstrated a role for the risk enhancer in controlling a network of genes, the mechanisms of the SNPs within enhancers remain to be determined. We attempted to create single-base-pair edits to generate isogenic cell lines with different genotypes of rs2737004. However, due to evidence of confounding effects on gene expression (possibly off-target edits), we did not move forward with that portion of the study. This limitation is worth noting because of the challenges related to technical constraints to single-base-pair-editing, requiring a PAM sequence near the targeted nucleotide. The choice of guide sequences is much more restricted which increases the chances for off-target edits. Future follow-up studies using cell lines with germline variation, similar to the methods of Langston et al.,<sup>44</sup> may be a better approach.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODELS AND STUDY PARTICIPANT DETAILS
  - iPSC cell lines
- METHODS DETAILS
  - iPSC differentiation to HPCs
  - Flow analysis of HPC markers
  - Differentiation of CD43<sup>+</sup> HPCs to microglia
  - Immunocytochemistry and imaging microglia
  - ATAC-seq
  - Sequencing of ATAC-seq libraries

- Identification of ATAC-seq peaks
- PD risk SNPs ATAC-seq peak intersect
- CRISPR/Cas9 deletion in iPSCs
- RNA-sequencing
- Sequencing of total RNA-seq libraries
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2024.108921>.

## ACKNOWLEDGMENTS

Van Andel Institute (VAI) provided the facilities and general funding support to conduct this study. This work was also supported in part by VAI core facilities. We would like to thank the Genomics Core (RRID:SCR\_022913), led by Marie Adams, for conducting and advising on next-generation sequencing. Rachael Sheridan (Flow Cytometry Core manager), Maddie Nichols, and Josh Schipper conducted flow analyses (RRID:SCR\_022685). Zach Madaj in the Bioinformatics and Biostatistics core provided support for statistical methods and analysis. Lastly, we would like to thank the Optical Imaging Core (RRID:SCR\_021968), led by Corine Esquibel. Lorna Cohen performed confocal microscopy, and Kristin Gallik curated final fluorescent images.

## AUTHOR CONTRIBUTIONS

A.B.: methodology, formal analysis, investigation, writing – original draft.  
S.E.P.: validation, writing – review and editing.  
E.J.C.v.d.S.: investigation, writing – review and editing.  
G.A.C.: conceptualization, supervision, funding acquisition, writing – review and editing.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 17, 2023

Revised: December 5, 2023

Accepted: January 12, 2024

Published: January 17, 2024

## REFERENCES

1. Nalls, M.A., Blauwendraat, C., Vallerga, C.L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D.A., Noyce, A.J., Xue, A., et al. (2019). Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* *18*, 1091–1102.
2. Gallagher, M.D., and Chen-Plotkin, A.S. (2018). The Post-GWAS Era: From Association to Function. *Am. J. Hum. Genet.* *102*, 717–730.
3. Soldner, F., Stelzer, Y., Shivalila, C.S., Abraham, B.J., Latourelle, J.C., Barrasa, M.I., Goldmann, J., Myers, R.H., Young, R.A., and Jaenisch, R. (2016). Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature* *533*, 95–99.
4. Coetzee, S.G., Pierce, S., Brundin, P., Brundin, L., Hazelett, D.J., and Coetzee, G.A. (2016). Enrichment of risk SNPs in regulatory regions implicate diverse tissues in Parkinson's disease etiology. *Sci. Rep.* *6*, 30509.
5. Pierce, S.E., Booms, A., Prah, J., van der Schans, E.J.C., Tyson, T., and Coetzee, G.A. (2020). Post-GWAS knowledge gap: the how, where, and when. *NPJ Parkinsons Dis.* *6*, 23.
6. Ho, M.S. (2019). Microglia in Parkinson's Disease. *Adv. Exp. Med. Biol.* *1175*, 335–353.
7. Reynolds, R.H., Botia, J., Nalls, M.A., International Parkinson's Disease Genomics Consortium; System Genomics of Parkinson's Disease, Hardy, J., Gagliano Taliun, S.A., and Ryten, M. (2019). Moving beyond neurons: the role of cell type-specific gene regulation in Parkinson's disease heritability. *NPJ Parkinsons Dis* *5*, 6.
8. Andersen, M.S., Bandres-Ciga, S., Reynolds, R.H., Hardy, J., Ryten, M., Krohn, L., Gan-Or, Z., Holtman, I.R., and Pihlström, L.; International Parkinson's Disease Genomics Consortium (2021). Heritability Enrichment Implicates Microglia in Parkinson's Disease Pathogenesis. *Ann. Neurol.* *89*, 942–951.
9. Hazelett, D.J., Rhie, S.K., Gaddis, M., Yan, C., Lakeland, D.L., Coetzee, S.G., Ellipse/GAME-ON consortium; Practical consortium, Henderson, B.E., Noushmehr, H., et al. (2014). Comprehensive functional annotation of 77 prostate cancer risk loci. *PLoS Genet.* *10*, e1004102.
10. Degtyareva, A.O., Antontseva, E.V., and Merkulova, T.I. (2021). Regulatory SNPs: Altered Transcription Factor Binding Sites Implicated in Complex Traits and Diseases. *Int. J. Mol. Sci.* *22*, 6454.
11. Gosselin, D., Skola, D., Coufal, N.G., Holtman, I.R., Schlachetzki, J.C.M., Sajti, E., Jaeger, B.N., O'Connor, C., Fitzpatrick, C., Pasillas, M.P., et al. (2017). An environment-dependent transcriptional network specifies human microglia identity. *Science* *356*, eaal3222.
12. Booms, A., Pierce, S.E., and Coetzee, G.A. (2020). Parkinson's disease genetic risk evaluation in microglia highlights autophagy and lysosomal genes. Preprint at bioRxiv.
13. Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., et al. (2014). Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* *46*, 989–993.
14. Pihlström, L., Blauwendraat, C., Cappelletti, C., Berge-Seidl, V., Langmyhr, M., Henriksen, S.P., van de Berg, W.D.J., Gibbs, J.R., Cookson, M.R., International Parkinson Disease Genomics Consortium, et al.. (2018). A comprehensive analysis of SNCA-related genetic risk in sporadic parkinson disease. *Ann. Neurol.* *84*, 117–129.
15. McQuade, A., Coburn, M., Tu, C.H., Hasselmann, J., Davtyan, H., and Blurton-Jones, M. (2018). Development and validation of a simplified method to generate

- human microglia from pluripotent stem cells. *Mol. Neurodegener.* 13, 67.
16. Johnston, A.D., Simões-Pires, C.A., Thompson, T.V., Suzuki, M., and Greally, J.M. (2019). Functional genetic variants can mediate their regulatory effects through alteration of transcription factor binding. *Nat. Commun.* 10, 3472.
  17. Nalls, M.A., Blauwendraat, C., Vallerga, C.L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D.A., Noyce, A.J., Xue, A., et al. (2019). Expanding Parkinson's disease genetics: novel risk loci, genomic context, causal insights and heritable risk. Preprint at bioRxiv. <https://doi.org/10.1101/388165>.
  18. Coetzee, S.G., Coetzee, G.A., and Hazelett, D.J. (2015). motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* 31, 3847–3849.
  19. Qin, T., Lee, C., Li, S., Cavalcante, R.G., Orchard, P., Yao, H., Zhang, H., Wang, S., Patil, S., Boyle, A.P., and Sartor, M.A. (2022). Comprehensive enhancer-target gene assignments improve gene set level interpretation of genome-wide regulatory data. *Genome Biol.* 23, 105.
  20. Nott, A., Holtman, I.R., Coufal, N.G., Schlachetzki, J.C.M., Yu, M., Hu, R., Han, C.Z., Pena, M., Xiao, J., Wu, Y., et al. (2019). Brain cell type-specific enhancer-promoter interactome maps and disease risk association. *Science* 366, 1134–1139.
  21. Hammal, F., de Langen, P., Bergon, A., Lopez, F., and Ballester, B. (2022). ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res.* 50, D316–D325.
  22. Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* 19, 789–800.
  23. Hansen, A.S. (2020). CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus* 11, 132–148.
  24. Pulido-Salgado, M., Vidal-Taboada, J.M., and Saura, J. (2015). C/EBPbeta and C/EBPdelta transcription factors: Basic biology and roles in the CNS. *Prog. Neurobiol.* 132, 1–33.
  25. Gómez-Santos, C., Barrachina, M., Giménez-Xavier, P., Dalfó, E., Ferrer, I., and Ambrosio, S. (2005). Induction of C/EBP beta and GADD153 expression by dopamine in human neuroblastoma cells. Relationship with alpha-synuclein increase and cell damage. *Brain Res. Bull.* 65, 87–95.
  26. Prah, J., Pierce, S.E., van der Schans, E.J., Coetzee, G.A., and Tyson, T. (2021). Global Effects of a PD Risk-SNP at the Alpha-Synuclein Locus. Preprint at bioRxiv. 2021. 2007.2006.451330.
  27. Dekker, J., Belmont, A.S., Guttman, M., Leshyk, V.O., Lis, J.T., Lomvardas, S., Mirny, L.A., O'Shea, C.C., Park, P.J., Ren, B., et al. (2017). The 4D nucleome project. *Nature* 549, 219–226.
  28. Ruan, C., and Elyaman, W. (2022). A New Understanding of TMEM119 as a Marker of Microglia. *Front. Cell. Neurosci.* 16, 902372.
  29. Pomaznoy, M., Ha, B., and Peters, B. (2018). GOnet: a tool for interactive Gene Ontology analysis. *BMC Bioinf.* 19, 470.
  30. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29.
  31. Blauwendraat, C., Nalls, M.A., and Singleton, A.B. (2020). The genetic architecture of Parkinson's disease. *Lancet Neurol.* 19, 170–178.
  32. Taguchi, K., Watanabe, Y., Tsujimura, A., and Tanaka, M. (2019). Expression of alpha-synuclein is regulated in a neuronal cell type-dependent manner. *Anat. Sci. Int.* 94, 11–22.
  33. Kasen, A., Houck, C., Burmeister, A.R., Sha, Q., Brundin, L., and Brundin, P. (2022). Upregulation of alpha-synuclein following immune activation: Possible trigger of Parkinson's disease. *Neurobiol. Dis.* 166, 105654.
  34. Calabresi, P., Di Lazzaro, G., Marino, G., Campanelli, F., and Ghiglieri, V. (2023). Advances in understanding the function of alpha-synuclein: implications for Parkinson's disease. *Brain* 146, 3587–3597.
  35. Posner, M.G. (2022). Multimerin-1 and cancer: a review. *Biosci. Rep.* 42.
  36. Yao, S., Zhang, X., Zou, S.C., Zhu, Y., Li, B., Kuang, W.P., Guo, Y., Li, X.S., Li, L., and Wang, X.Y. (2021). A transcriptome-wide association study identifies susceptibility genes for Parkinson's disease. *NPJ Parkinsons Dis.* 7, 79.
  37. Ferese, R., Modugno, N., Campopiano, R., Santilli, M., Zampatti, S., Giardina, E., Nardone, A., Postorivo, D., Fornai, F., Novelli, G., et al. (2015). Four Copies of SNCA Responsible for Autosomal Dominant Parkinson's Disease in Two Italian Siblings. *Parkinsons Dis.* 2015, 546462.
  38. Stern, S., Lau, S., Manole, A., Rosh, I., Percia, M.M., Ben Ezer, R., Shokhirev, M.N., Qiu, F., Schafer, S., Mansour, A.A., et al. (2022). Reduced synaptic activity and dysregulated extracellular matrix pathways in midbrain neurons from Parkinson's disease patients. *NPJ Parkinsons Dis.* 8, 103.
  39. Lauro, C., and Limatola, C. (2020). Metabolic Reprogramming of Microglia in the Regulation of the Innate Inflammatory Response. *Front. Immunol.* 11, 493.
  40. Lang, L., Chennalakuzhy, R., Shay, C., and Teng, Y. (2019). PFKF Signaling at a Glance: An Emerging Mediator of Cancer Cell Metabolism. *Adv. Exp. Med. Biol.* 1134, 243–258.
  41. Kaplon, J., van Dam, L., and Peeper, D. (2015). Two-way communication between the metabolic and cell cycle machineries: the molecular basis. *Cell Cycle* 14, 2022–2032.
  42. Edison, P., Ahmed, I., Fan, Z., Hinz, R., Gelosa, G., Ray Chaudhuri, K., Walker, Z., Turkheimer, F.E., and Brooks, D.J. (2013). Microglia, amyloid, and glucose metabolism in Parkinson's disease with and without dementia. *Neuropsychopharmacology* 38, 938–949.
  43. De Pablo-Fernandez, E., Goldacre, R., Pakpoor, J., Noyce, A.J., and Warner, T.T. (2018). Association between diabetes and subsequent Parkinson disease: A record-linkage cohort study. *Neurology* 91, e139–e142.
  44. Langston, R.G., Beilina, A., Reed, X., Kaganovich, A., Singleton, A.B., Blauwendraat, C., Gibbs, J.R., and Cookson, M.R. (2022). Association of a common genetic variant with Parkinson's disease is mediated by microglia. *Sci. Transl. Med.* 14, eabp8869.
  45. Deng, D., and Yan, N. (2016). GLUT, SGLT, and SWEET: Structural and mechanistic investigations of the glucose transporters. *Protein Sci.* 25, 546–558.
  46. Gluchowska, K., Pliszka, M., and Szablewski, L. (2021). Expression of glucose transporters in human neurodegenerative diseases. *Biochem. Biophys. Res. Commun.* 540, 8–15.
  47. Siddiqui, I.J., Pervaiz, N., and Abbasi, A.A. (2016). The Parkinson Disease gene SNCA: Evolutionary and structural insights with pathological implication. *Sci. Rep.* 6, 24475.
  48. Yamasaki, R., Lu, H., Butovsky, O., Ohno, N., Rietsch, A.M., Cialic, R., Wu, P.M., Doynkan, C.E., Lin, J., Cotleur, A.C., et al. (2014). Differential roles of microglia and monocytes in the inflamed central nervous system. *J. Exp. Med.* 211, 1533–1549.
  49. Grozdanov, V., and Danzer, K.M. (2020). Intracellular Alpha-Synuclein and Immune Cell Function. *Front. Cell Dev. Biol.* 8, 562692.
  50. Ying, Z.M., Lv, Q.K., Yao, X.Y., Dong, A.Q., Yang, Y.P., Cao, Y.L., Wang, F., Gong, A.P., and Liu, C.F. (2022). BAG3 promotes autophagy and suppresses NLRP3 inflammasome activation in Parkinson's disease. *Ann. Transl. Med.* 10, 1218.
  51. Zhao, J., Wei, J., Mialki, R.K., Mallampalli, D.F., Chen, B.B., Coon, T., Zou, C., Mallampalli, R.K., and Zhao, Y. (2012). F-box protein FBXL19-mediated ubiquitination and degradation of the receptor for IL-33 limits pulmonary inflammation. *Nat. Immunol.* 13, 651–658.
  52. Wei, J., Mialki, R.K., Dong, S., Khoo, A., Mallampalli, R.K., Zhao, Y., and Zhao, J. (2013). A new mechanism of RhoA ubiquitination and degradation: roles of SCF(FBXL19) E3 ligase and Erk2. *Biochim. Biophys. Acta* 1833, 2757–2764.
  53. Socodato, R., Portugal, C.C., Canedo, T., Rodrigues, A., Almeida, T.O., Henriques, J.F., Vaz, S.H., Magalhães, J., Silva, C.M., Baptista, F.I., et al. (2020). Microglia Dysfunction Caused by the Loss of RhoA Disrupts Neuronal Physiology and Leads to Neurodegeneration. *Cell Rep.* 31, 107796.
  54. Wang, S., Bleack, A., Nadif Kasri, N., Kleefstra, T., van Rhijn, J.R., and Schubert, D. (2021). SETD1A Mediated H3K4 Methylation and Its Role in Neurodevelopmental and Neuropsychiatric Disorders. *Front. Mol. Neurosci.* 14, 772000.
  55. Chong, Z.S., Khong, Z.J., Tay, S.H., and Ng, S.Y. (2022). Metabolic contributions to neuronal deficits caused by genomic disruption of schizophrenia risk gene SETD1A. *Schizophrenia (Heidelberg)* 8, 115.
  56. Spataro, N., Calafell, F., Cervera-Carles, L., Casals, F., Pagonabarraga, J., Pascual-Sedano, B., Campolongo, A., Kulisevsky, J., Lleó, A., Navarro, A., et al. (2015). Mendelian genes for Parkinson's disease contribute to the sporadic forms of the disease. *Hum. Mol. Genet.* 24, 2023–2034.
  57. Huang, P., Cai, Y., Zhao, B., and Cui, L. (2018). Roles of NUCKS1 in Diseases: Susceptibility, Potential Biomarker, and Regulatory Mechanisms. *BioMed Res. Int.* 2018, 7969068.
  58. Singh, S., and Seth, P.K. (2019). Functional association between NUCKS1 gene and Parkinson disease: A potential susceptibility biomarker. *Bioinformation* 15, 548–556.
  59. Mazo, G. (2021). QuickFigures: A toolkit and ImageJ Plugin to quickly transform microscope images into scientific figures. *PLoS One* 16, e0240280.
  60. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
  61. Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples

- in a single report. *Bioinformatics* 32, 3047–3048.
62. Faust, G.G., and Hall, I.M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* 30, 2503–2505.
  63. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
  64. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.
  65. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9, e1003118.
  66. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
  67. Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383.
  68. Zheng, Q., Cai, X., Tan, M.H., Schaffert, S., Arnold, C.P., Gong, X., Chen, C.Z., and Huang, S. (2014). Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. *Biotechniques* 57, 115–124.
  69. Liu, Y., Zhou, J., and White, K.P. (2014). RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* 30, 301–304.
  70. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
  71. Nikolayeva, O., and Robinson, M.D. (2014). edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. *Methods Mol. Biol.* 1150, 45–79.
  72. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308.
  73. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, giab008.
  74. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.
  75. Wang, Q., Li, M., Wu, T., Zhan, L., Li, L., Chen, M., Xie, W., Xie, Z., Hu, E., Xu, S., et al. (2022). Exploring epigenomic datasets by ChIPseeker. *Curr. Protoc.* 2, e585.
  76. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
  77. Nassar, L.R., Barber, G.P., Benet-Pages, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J.N., Hinrichs, A.S., Lee, B.T., et al. (2023). The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res.* 51, D1188–D1195.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Mouse APC anti-human CD43	BioLegend	cat#343205; RRID: <a href="#">AB_2194072</a>
Mouse anti-human Iba1	Thermo Fisher Scientific	GT10312; cat#MA5-27726; RRID: <a href="#">AB_2735228</a>
Alexa Fluor 488 goat anti-mouse	Thermo Fisher Scientific	cat#A-11001; RRID: <a href="#">AB_2534069</a>
Rabbit anti-human TMEM119	Thermo Fisher Scientific	cat#PA5-119902; RRID: <a href="#">AB_2913474</a>
Alexa Fluor 594 goat anti-rabbit	Thermo Fisher Scientific	cat#A-11012; RRID: <a href="#">AB_2534079</a>
<b>Chemicals, peptides, and recombinant proteins</b>		
insulin-transferrin-selenium (100X)	Thermo Fisher Scientific	cat#41400045
B27	Thermo Fisher Scientific	cat#17504044
N2	Thermo Fisher Scientific	cat#17502048
glutamax	Thermo Fisher Scientific	cat#35050061
non-essential amino acids	Thermo Fisher Scientific	cat#11140050
monothioglycerol	Sigma	cat#A8960-5G
human insulin	Sigma	cat#I2643-50MG
IL-34	Peprotech	cat#200-34
TGFβ1	Peprotech	cat#100-21
M-CSF	Peprotech	cat#300-25
CD200	Novoprotein	cat#C311
CX3CL1	Peprotech	cat#300-31
Poly(ethyleneimine) solution (PEI)	Milipore Sigma	cat#181978-100g
4% paraformaldehyde	Electron Microscopy Sciences	cat#157-8-100
DPBS	Gibco	cat#14190-144
Goat serum	Abcam	cat#ab7481
Triton-x	Fisher Scientific	cat#AAA16046AE
NucBlue fixed stain ready probes	Thermo Fisher Scientific	cat#R37606
Poly-D-Lysine	Gibco	cat#A38904-10
Tris-HCl pH 7.5	Thermo Fisher Scientific	cat#15567-027
NaCl	Fisher	cat#S271-500
MgCl <sub>2</sub>	Fisher	cat#bp214-500
nuclease free H <sub>2</sub> O	Invitrogen	cat#AM9938
NP-40	Thermo Fisher Scientific	cat#85124
Tween-20	Fisher	cat#bp337-500
Digitonin	Promega	cat#G9441
<b>Critical commercial assays</b>		
hematopoietic kit	STEMdiff	cat#05310
Nextera DNA library prep kit	Illumina	cat#FC-121-1030
Zymo clean and concentrator kit	Zymo	cat# D4014
QuantiFluor® dsDNA System	Promega	cat#E2671
NextSeq 500/550 150 bp sequencing kit (v2)	Illumina	cat# 20024907
Lonza Human Stem Cell Nucleofactor Kit 1	Lonza	cat#VPH-5012
QIAshredder	Qiagen	cat#79654

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RNeasy isolation kit	Qiagen	cat#74104
Takara SMARTer Stranded Total RNA-Seq Kit v3 Pico Input Mammalian	Takara Bio	cat#634485
QIAprep Spin Miniprep Kit (250)	Qiagen	cat#27106
QIAquick Gel extraction kit	Qiagen	cat# 28704/28706

**Deposited data**

raw and analyzed data	this paper	GEO: <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE245524">GSE245524</a>
reanalyzed microglia ATAC-seq and ChIP-seq data	Gosselin et al. <sup>11</sup>	dbGAP, accession number: phs001373.v1.p1
reanalyzed PLAC-seq data	Nott et al. <sup>20</sup>	dbGAP, accession number: phs001373.v2.p1

**Experimental models: Cell lines**

Human iPSCs	ATCC-DYS0100	cat#ACS-1019; RRID:CVCL_X499
Human iPSCs	Synthego	PGP1-SV1

**Oligonucleotides**

Illumina primer 1 AATGATACGGCGACCACCGA	IDT	N/A
Illumina primer 2 CAAGCAGAAGACGGCATAACGA	IDT	N/A
SNCA_guide_1_F caccgGTGAAGGTATCCGTATAATG	this paper	N/A
SNCA_guide_1_R aaacCATTATACGGATACCTTCACc	this paper	N/A
SNCA_guide_2_F caccgCAATGACTTTCGGTACTG	this paper	N/A
SNCA_guide_2_R aaacCAGTATAccGAAAGTCATTGc	this paper	N/A

**Recombinant DNA**

pSpCas9(BB)-2A-GFP (PX458)	Addgene Ran et al. <sup>72</sup>	cat#48138; RRID:Addgene_48138
----------------------------	-------------------------------------	-------------------------------

**Software and algorithms**

Illumina NextSeq Control Software (NCS) v2.0	Illumina	<a href="https://support.illumina.com/sequencing/sequencing_instruments/nextseq-500/downloads.html">https://support.illumina.com/sequencing/sequencing_instruments/nextseq-500/downloads.html</a>
Illumina Bcl2fastq v1.9.0	Illumina	<a href="https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software/downloads.html">https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software/downloads.html</a>
Trimgalore	Felix Krueger, Babraham Institute	<a href="https://github.com/FelixKrueger/TrimGalore">https://github.com/FelixKrueger/TrimGalore</a>
BWA v0.7.17	Li et al. <sup>60</sup>	<a href="https://anaconda.org/bioconda/bwa">https://anaconda.org/bioconda/bwa</a> ; RRID:SCR_010910
Multiqc v1.0	Ewels et al. <sup>61</sup>	<a href="https://anaconda.org/bioconda/multiqc">https://anaconda.org/bioconda/multiqc</a> ; RRID:SCR_014982
Samblaster v0.1.24	Faust et al. <sup>62</sup>	<a href="https://anaconda.org/bioconda/samblaster">https://anaconda.org/bioconda/samblaster</a> ; RRID:SCR_000468
Samtools v1.9	Danecek et al. <sup>73</sup>	<a href="https://anaconda.org/bioconda/samtools">https://anaconda.org/bioconda/samtools</a> ; RRID:SCR_002105
MACS2 v2.1.1	Zhang et al. <sup>64</sup>	<a href="https://anaconda.org/bioconda/macs2">https://anaconda.org/bioconda/macs2</a> ; RRID:SCR_013291

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GenomicRanges v3.11	Lawrence et al. <sup>65</sup>	<a href="https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html">https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html</a> ; RRID:SCR_000025
Bedtools v2.29.0	Quinlan et al. <sup>66</sup>	<a href="https://anaconda.org/bioconda/bedtools">https://anaconda.org/bioconda/bedtools</a> ; RRID:SCR_006646
Benchling	N/A	<a href="https://www.benchling.com/RRID:SCR_013955">https://www.benchling.com/RRID:SCR_013955</a>
IGV v2.16.2	Robinson et al. <sup>74</sup>	<a href="https://igv.org/doc/desktop/#DownloadPage/">https://igv.org/doc/desktop/#DownloadPage/</a> ; RRID:SCR_011793
ChIPseeker v3.11	Wang et al. <sup>75</sup> Yu et al. <sup>67</sup>	<a href="https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html">https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html</a> ; RRID:SCR_021322
STAR v2.5.4b	Dobin et al. <sup>70</sup>	<a href="https://anaconda.org/bioconda/star">https://anaconda.org/bioconda/star</a> ;
edgeR v3.18	Robinson et al. <sup>76</sup>	<a href="https://bioconductor.org/packages/release/bioc/html/edgeR.html">https://bioconductor.org/packages/release/bioc/html/edgeR.html</a> ; RRID:SCR_012802
Graphpad Prism v10.0.3	N/A	<a href="https://www.graphpad.com">https://www.graphpad.com</a> ; RRID:SCR_002798
UCSC Genome Browser PLAC-seq, H3K27ac ChIP-seq, and ATAC-seq session	Nott et al. <sup>20</sup>	<a href="https://genome.ucsc.edu/s/nottalex/glassLab_BrainCellTypes_hg19">https://genome.ucsc.edu/s/nottalex/glassLab_BrainCellTypes_hg19</a>
UCSC Genome Browser	Nassar et al. <sup>77</sup>	<a href="https://genome.ucsc.edu">https://genome.ucsc.edu</a> ; RRID:SCR_005780

**Other**

StemFlex medium	ThermoFisher	cat#A3349401
Geltrex LDEV-free reduced growth factor basement membrane	ThermoFisher	cat#A1413201
iMatrix	Matrixome	cat#892012
ReLeSR	STEMCELL Technologies	cat#05872
Cell Staining Buffer	BioLegend	cat#420201
TruStain FcX (Fc Receptor Blocking Solution)	BioLegend	cat#422301
12x75mm round bottom tubes	Fisherbrand	cat#14-965-3C
DMEM/F12, HEPES, no phenol red	ThermoFisher	cat#11039021
8-well chamber slides	Ibidi	cat#80841
EverBright Hardset Mounting Medium	Biotium	cat#23003
TD Buffer (part of kit)	Illumina	cat#FC-121-1030
ATM (part of kit)	Illumina	cat#FC-121-1030
NT buffer (part of kit)	Illumina	cat#FC-121-1030
Illumina Nextera DNA unique Dual Indexes	Illumina	cat#20027214
NEBNext High-Fidelity 2X PCR Master Mix	NEB	cat#M0541L
KAPA Pure beads	Roche	cat#KK8001
Agilent DNA High Sensitivity chip	Agilent Technologies	cat#5067-4626
T4 Kinase (PNK)	Thermo Fisher Scientific	cat#EK3001
BbsI	NEB	cat#R0539S/R0539L
Cut Smart Buffer	NEB	cat#B7204
T4 ligase buffer	Thermo Fisher Scientific	cat#B69
T4 DNA ligase	Thermo Fisher Scientific	cat# EL0011
PlasmidSafe ATP-dependent DNase	Lucigen	cat#E3101K
PlasmidSafe buffer	Lucigen	not available
25 mM ATP solution	Lucigen	not available
One Shot Top10 E. Coli	Thermo Fisher Scientific	cat#C404003
S.O.C.	Thermo Fisher Scientific	cat#15544034
Revitacell	Thermo Fisher Scientific	cat#A2644501

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the [lead contact](#), Alix Booms ([alix.booms@vai.org](mailto:alix.booms@vai.org)).

### Materials availability

There are restrictions to the availability of edited PGP1 cell lines due to the requirement of a materials transfer agreement (MTA).

### Data and code availability

- Bulk RNA-seq and ATAC-seq data have been deposited at GEO and are publicly available as of the date of publication. This paper also analyzes existing publicly available data. Accession numbers are listed in the [key resources table](#).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODELS AND STUDY PARTICIPANT DETAILS

### iPSC cell lines

For ATAC-seq experiments, induced pluripotent stem cells (iPSCs) were obtained from ATCC (ACS-1019, DYS0100, male neonate). For CRISPR editing experiments, iPSCs were obtained from Synthego (PGP1-SV1, Male age 55). All validation and QC of cell lines were performed by the supplier. We have not authenticated these cells following receipt. Conditions for culturing both cell lines include 5% CO<sub>2</sub> at 37°C. At the iPSC stage, cells were cultured in StemFlex medium on either Geltrex LDEV-free reduced growth factor basement membrane for the ATCC cell line or iMatrix for the PGP1 cell line. When cells reached 80% confluency, they were passaged using ReLeSR.

## METHODS DETAILS

### iPSC differentiation to HPCs

iPSCs were first differentiated to hematopoietic progenitors (microglia precursors) using the STEMdiff hematopoietic kit per the methods detailed by McQuade et al.<sup>15</sup> On day -1 iPSCs are seeded in 6-well plates and allowed to adhere overnight. On day 0, StemFlex is removed and replaced with medium A. On day 2 half of the medium was changed out with fresh medium A. On day 3 all media was changed to medium B. Half medium changes were then done on days 5, 7, and 10. Cells were harvested on day 12 and assessed for CD43 expression using Flow.

### Flow analysis of HPC markers

To confirm iPSC differentiation to HPCs on day 12, cells were stained for CD43. We followed the "Cell Surface Flow Cytometry Staining Protocol" from Biolegend. Supernatant and non-adherent cells were collected from each well of a 6-well plate. They were then pelleted and re-suspended in 5 mL for counting. Single suspensions of 200,000 cells were prepared in up to 15 mL of Cell Staining Buffer. Cells were centrifuged at 350 × g for 5 minutes and supernatant was discarded. The pellet was then resuspended in Cell Staining Buffer (100 µl/# of conditions). Five microliters of TruStain FcX (Fc Receptor Blocking Solution) was added to each sample followed by an incubation at room temperature for 5-10 minutes. After incubation, 200,000 cells per condition were aliquoted into culture test tubes (Fisherbrand 12 × 75mm), one for CD43 and one for the non-stained control. Five microliters of CD43 was then added to one sample and allowed to incubate for 15-20 min in the dark. Cells were washed two times with at least 3 mL of Cell Staining Buffer. At the final wash the pellet was resuspended in 300 µl of Cell Staining Buffer plus 10.9 mM DAPI at a concentration of 3µM. Using a Beckman Coulter CytoFLEX S Flow cytometer. We determined that all cultures were pure if over 90% of cells assayed were positive for CD43 (per McQuade et al.<sup>15</sup>). See [Figure S3B](#) and [S5 A](#) for Flow results on the ATCC and PGP1 cell lines respectively.

### Differentiation of CD43<sup>+</sup> HPCs to microglia

On day 0, HPCs were plated at 100,000 cells per 6-well plate on iMatrix in 2 mL of microglia medium: DMEM/F12, 2X insulin-transferrin-selenium, 2X B27, 0.5X N2, 1X glutamax, 1x non-essential amino acids, 400 µM monothioglycerol, 5 µg/mL insulin. This media was supplemented with 100 ng/mL IL-34, 50 ng/mL TGFβ, and 25 ng/mL M-CSF. On days 2, 4, 6, 8, and 10, 1 mL of microglia medium plus 3 freshly thawed cytokines were added to each well. On day 12, all but 1 mL was collected from each well and spun down at 300 rcf for 5 minutes. The pellet was resuspended in 1 mL/well fresh medium plus 3 cytokines and added back to the same plate. Media was supplemented again with 1 mL/well fresh microglia media plus 3 cytokines on days 14, 16, 18, 20, 22, and 24. On day 25 all but 1 mL was removed from each well and spun down at 300 rcf for 5 minutes. Cells were then resuspended in 1 mL/well fresh microglia medium plus 5 cytokines (100 ng/mL IL-34, 50 ng/mL TGFβ, 25 ng/mL M-CSF, 100 ng/mL CD200 and 100 ng/mL CX3CL1). Cells were collected on day 28 for ICC and RNA-seq.



### Immunocytochemistry and imaging microglia

Following differentiation of HPCs to microglia, expression of microglia-specific markers were confirmed using a mouse anti-human Iba1 primary antibody with Alexa Fluor 488 goat anti-mouse secondary antibody. We also used a rabbit anti-human TMEM119 primary antibody with Alexa Fluor 594 goat anti-rabbit secondary antibody.

The staining procedure differed for each cell line. For the ATCC cell line, cells were plated in 24-well plates on glass coverslips coated with PEI and allowed to adhere for 24-hours in the incubator (5% CO<sub>2</sub> at 37°C). They were then fixed with 4% paraformaldehyde and permeabilized (DPBS, goat serum, and triton-x). Cells were incubated in blocking buffer (DPBS and goat serum) with primary antibodies overnight at 4°C. After 24-hours, cells were washed with DPBS and incubated with secondary antibodies in DPBS for 30 minutes. They were then washed again prior to adding NucBlue Fixed stain ready probes. Coverslips were removed and mounted on glass slides using EverBright Hardset Mounting Medium. For PGP1 cell lines, staining was done the same as the ATCC cell lines, except they were plated on Poly-D-Lysine in Ibidi 8-well chamber slides and allowed to adhere at 37°C for 2 hours prior to fixing. For each cell line imaging was done as follows.

ATCC: Fluorescent images for [Figure S5 A](#) were taken with the Nikon Eclipse microscope with NIS Elements software (version 5.11.01) and exported as a tiffs. Bright field images were taken directly in culture plates using the EVOS microscope. The final panel of images were compiled in Power Point.

PGP1 wild-type and edited: Confocal Z-stacks for [Figure S5 B](#) were collected using a Zeiss LSM 880 equipped with an Axio Observer 7 inverted microscope body and acquired with Zen Black (version 2.3) software using 405nm diode, 488nm argon ion, and 561nm DPSS laser lines. Emitted light was detected through a Zeiss Plan-apochromat 63x/1.4 NA oil immersion objective, using an Airyscan GaAsP detector. Images were collected sequentially in 1024x1024 pixel resolution, using 0.25um z-steps. Images were acquired with an optical zoom of either 1.0 or 2.0, and individual voxels were therefore 0.13x0.13x0.25um or 0.07x0.07x0.25um (xyz), respectively.

To create [Figure S5B](#), raw czi images were opened in Fiji ImageJ (v1.54f). An average intensity z-projection was generated for each image, inclusive of all channels and z, and saved as a tiff. Any image larger than 1012x1012 pixels were cropped for uniformity. All six average intensity projection images were then concatenated for easier import into the ImageJ plugin QuickFigures<sup>59</sup> to assemble into the final figure shown.

### ATAC-seq

The biostatistics core at VARI conducted a power calculation based on ATAC-seq effect size to determine the number of optimal replicates. With three replicates and an average depth of ~40 reads per million, this study has >80% power to detect, with 95% confidence, peaks with ~1.8 fold or greater difference in accessibility and ~99% power for a ~2-fold difference in accessibility. This calculation was done for the purpose of performing a differential accessibility analysis between wild-type and edited microglia in future experiments. We believe the 4 replicates that we generated in combination with the 13 published ATAC-seq data is sufficient to detect robust peaks in microglia.

Microglia were thawed and cultured for at least one week prior to an ATAC-seq experiment. Samples that yielded the best fragmentation started from a total of 10K, 31K, and 100K cells. The pre-specified number of cells were aliquoted into 1.5 mL tubes and centrifuged at 400 x g for 7 minutes. The supernatant was removed, and the cells were washed once with 50 µl ice-cold PBS. The cells were then resuspended in ice-cold Lysis Buffer containing resuspension buffer (1M Tris-HCl pH 7.5 (final conc. = 10mM), 5 M NaCl (final conc. = 10 mM), 1M MgCl<sub>2</sub> (final conc. = 3 mM), and nuclease-free H<sub>2</sub>O), 10% NP-40 (final conc. = 0.1% v/v), 10% Tween-20 (final conc. = 0.1% v/v), and 1% Digitonin (final conc. = 0.01% v/v). Cells were then incubated on ice for 3 minutes. One mL of wash buffer (990 µl resuspension buffer + 10 µl Tween-20 (final conc. = 0.01% v/v)) was added to each tube. The tubes were then inverted 3X gently and centrifuged at 500 x g for 5 minutes. For each sample, 10 µl of transposition mix (7.5 µl 2X TD Buffer, 2.05 µl 1X PBS, 0.15 µl 10% Tween-20 (final conc. = 0.1 v/v), 1% Digitonin (final conc. = 0.01% v/v), and 0.15 nuclease-free H<sub>2</sub>O) was added. Five µl of ATM was then added separately to each sample. The samples were incubated for 60 minutes on a thermomixer at 1,000 rpm. Following incubation, the samples were placed on ice, and 5 µl of NT buffer was added to each tube to neutralize the tagmentation reaction. Tubes were then centrifuged at 300 x g at 20°C for 1 minute and incubated at room temp for 5 minutes. DNA purification was done using the Zymo clean and concentrator kit.

For library generation, 5 µl of Illumina Nextera DNA unique Dual Indexes plus 25 µl NEBNext High-Fidelity 2X PCR Master Mix was added to 20 µl of purified transposed DNA. The transposed fragments were amplified starting at 72°C for 5 minutes, 98°C for 30 seconds and then five cycles of 98°C for 10 seconds, 63°C for 30 seconds, and 72°C for 1 minute. qPCR was used to determine how many additional cycles to run on each sample. The PCR mix was composed of 5 µl of the partially amplified library from the previous step, 0.5 µl Illumina primer 1 (25 µM), 0.5 µl Illumina primer 2 (25 µM) 0.75 µl 20X Eva Green, and 5 µl NEBNext High-Fidelity 2X PCR Master Mix. Cycle conditions were set to 98°C for 30 seconds, and 20 cycles of 98°C for 10 seconds, 63°C for 30 seconds, and 72°C for 1 minute. The R vs. cycle number was plotted on a linear scale. Additional cycles were calculated by determining the number of cycles needed to reach 1/3 of the maximum R. PCR was continued on the remaining partially amplified libraries for the appropriate number of cycles calculated in the previous step.

### Sequencing of ATAC-seq libraries

Library quantification, size selection, and sequencing were carried out by the Genomics Core at VAI. PCR amplified libraries were size selected for fragments 200-800 bp in length using double-sided SPRI selection (0.5x followed by 1x) with KAPA Pure beads. The quality and quantity of the finished libraries were assessed using a combination of Agilent DNA High Sensitivity chip (Agilent Technologies, Inc.), and QuantiFluor® dsDNA System. Seventy five base pair, paired-end sequencing was performed on an Illumina NextSeq 500 sequencer using a 150 bp sequencing kit (v2) to produce a minimum of 50M paired-reads per library. Base-calling was done by Illumina NextSeq Control Software (NCS) v2.0, and the output of NCS was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v1.9.0.

### Identification of ATAC-seq peaks

Four replicates of one iPSC-derived microglia cell line and thirteen replicates of ATAC-seq data from different primary microglia samples (published data) were used to find consensus ATAC-seq peaks. ATAC-seq peak data from primary microglia were obtained from dbGAP deposited by the Glass lab (see [key resources table](#) for accession number).<sup>11</sup> All data from iPSC-derived and primary microglia were processed in the same way. The sequencing depth for iPSC-derived microglia was about 40-50 million, and the read length was about 75 base pairs. For primary microglia, samples were sequenced to a depth ranging from 20-50 million reads, and the read length ranged from 47-76 base pairs.<sup>11</sup> Using Trimgalore, reads were trimmed or removed if they were below 20 base pairs in length or had a quality score below 20. All other parameters were default. Forward and reverse reads for iPSC-derived microglia, and single-end reads for primary microglia were then aligned to the hg19 genome using the default settings for BWA v0.7.17.<sup>60</sup> Multiqc v1.0<sup>61</sup> was then run on all samples following alignment. Samblaster v0.1.24<sup>62</sup> was used to sort and mark duplicate reads in bam files, and Samtools v1.9<sup>63</sup> was used to remove duplicate reads and index the bam files. Peaks were then called using MACS2 v2.1.1<sup>64</sup> default parameters. GenomicRanges v3.11<sup>65</sup> was used to generate a consensus peak set (starting from narrowPeak files). We used peaks present in 3/4 samples for iPSC-derived microglia and peaks that were present in 10/13 primary microglia samples.

### PD risk SNPs ATAC-seq peak intersect

The list of 6,749 SNPs was obtained from the Nalls et al. bioRxiv version<sup>17</sup> from a supplemental file labeled "SNPs of interest tagging genes for functional inferences and networks analysis." Using Bedtools v2.29.0,<sup>66</sup> we searched for overlaps between the location of PD risk SNPs and ATAC-seq peaks from iPSC-derived and primary microglia. The intersecting regions were then evaluated in IGV v2.16.2 for SNPs that overlapped or were nearby (within 100 bp) of an H3K37ac ChIP-seq peaks from primary microglia. In this analysis, we found a total of 73 SNPs that we then ranked by GWAS p-value ([Table S1](#) and related [Table 1](#)). The locations of the 73 SNPs in ATAC-seq peaks were annotated using ChIPseeker v3.11.<sup>67</sup>

### CRISPR/Cas9 deletion in iPSCs

The PGP1 cell line was used to create edited cell lines. CRISPR guides were designed using Benchling. Guide sequences are as follows (lower-case letters denote the sequences added for ligation into the pSpCas9(BB)-2A-GFP (PX458) vector from Addgene. See [key resources table](#) for guide sequences. For CRISPR/Cas9 editing, we used methods published by the Zheng lab.<sup>68</sup> To generate plasmids, 1  $\mu$ l 100  $\mu$ M forward and 1  $\mu$ l 100  $\mu$ M reverse guide sequences were annealed and phosphorylated using 10X T4 Ligase Buffer and 0.5  $\mu$ l T4 Kinase (PNK) respectively. The reaction was brought up to 10  $\mu$ l with ddH<sub>2</sub>O and incubated in the thermal cycler at 37°C for 30 minutes and 95°C for 5 minutes with a ramp down to 25°C at 5°C/min and a hold 10°C. The donor plasmid was digested in a 20  $\mu$ l reaction with 1  $\mu$ l BbsI, 2  $\mu$ l PX458 vector, 2  $\mu$ l Cut Smart Buffer, and 15  $\mu$ l ddH<sub>2</sub>O. The reaction was then incubated at 37°C for 5-15 minutes. The vector was gel purified using a 0.8% agarose gel with SYBR safe and the QIAquick Gel extraction kit. Guide duplex sequences were ligated with the PX458 vector using 1  $\mu$ l diluted oligo duplex (1:250), 50 ng of digested vector, 2  $\mu$ l 10X T4 ligase buffer, 0.2  $\mu$ l T4 ligase (final amount = 1 Weiss U), and ddH<sub>2</sub>O in a final reaction volume of 20  $\mu$ l. The reaction was incubated at room temperature for 10 minutes. Following incubation, 0.4  $\mu$ l PlasmidSafe ATP-dependent DNase, 0.8  $\mu$ l 10 Mm ATP (25mM ATP soln.), and 2  $\mu$ l 10X PlasmidSafe Buffer was added to the 20  $\mu$ l ligation reaction and incubated at room temperature for 30 minutes.

The final vector was transformed using Top10 chemically competent E. coli. One to five  $\mu$ l of plasmid was added to single aliquot E. coli, incubated on ice for 30 minutes, heat shocked without shaking at 42°C, and placed on ice for 2 minutes. Pre-warmed S.O.C (250  $\mu$ l) was added prior to incubating vials horizontally at 37°C for 1 hour at 225 rpm in a shaking incubator. Cultures (25-100  $\mu$ l) were spread on pre-warmed ampicillin plates (125  $\mu$ g/mL) and incubated overnight at 37°C. A well isolated colony was inoculated into a culture of 1-5 mL LB containing ampicillin (100  $\mu$ g/mL) and incubated 37°C with vigorous shaking for 12-16 hours. Plasmids were isolated using QIAprep Spin Miniprep Kit (250).

For nucleofection of plasmids contain CRISPR guides we used Lonza Human Stem Cell Nucleofector Kit 1. Four  $\mu$ g total plasmid (2  $\mu$ g of each plasmid) was added to 100  $\mu$ l of nucleofector solution. The nucleofector solution plus plasmid was added to 400,000 pelleted iPSCs. Cells and solution were then added to the supplied cuvette and electroporated on program A23 using an Amaxa Nucleofector II. Cells were then immediately placed in 24-well plates with growth media plus revitacell. Following nucleofection, cells were allowed to grow for 48 hours prior to sorting. GFP-positive cells were sorted using a BD FACSymphony S6 cell sorter into 96-well plates (one cell per well) and expanded for at least 3 passages from the 96-well plate to a 24-well and finally up to a 6-well plate. Each clone was then PCR screened for the deletion ([Figure S4A](#)). We chose two separate clones named A11 and H11 (based on their position in the 96-well plate) to move forward with differentiation and RNA-sequencing.

### RNA-sequencing

RNA was collected using the QIAGEN QIAshredder and RNeasy isolation kit. Collection happened at three time-points across two separate differentiations. Sample names, their collection times, and total number of technical and biological reps can be found in [Figure S4B](#). We did not perform a power calculation prior to this experiment. However, Liu et al. demonstrates that the power (with FDR = 0.05) achieved by sequencing to a depth of 30M reads or above (we used 50M) with 3 or 4 replicates is about 0.75 and 0.85 respectively.<sup>69</sup> Paired-end reads

that had a quality score below 20 and were less than 20 base pairs in length were removed using Trimgalore (see [key resources table](#) for citation). Reads were then aligned with STAR<sup>70</sup> and quality checked using MulitQC.<sup>61</sup>

### Sequencing of total RNA-seq libraries

Libraries were prepared by the Van Andel Genomics Core from 10 ng of total RNA using Takara SMARTer Stranded Total RNA-Seq Kit v3 Pico Input Mammalian per the manufacturer's protocol. In brief, RNA was sheared to 300-400 bp, after which dscDNA was generated using a template switching mechanism, and unique dual indexed adapters were added to each sample. Ribosomal cDNA was degraded by scZapR and scrRNA probes, and libraries amplified with 13 cycles of PCR. Quality and quantity of the finished libraries were assessed using a combination of Agilent DNA High Sensitivity chip, QuantiFluor® dsDNA System, and Kapa Illumina Library Quantification qPCR assays. Individually indexed libraries were pooled and 100 bp, paired-end sequencing was performed on an Illumina NovaSeq6000 sequencer, to return a minimum read depth of 50M read pairs per library. Base calling was done by Illumina RTA3 and output of NCS was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v1.9.0

### QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis for differential gene expression in microglia was done using edgeR version 3.18.<sup>71</sup> Comparisons were made by grouping all technical and biological replicates of the edited microglia (A11 and H11, n = 4) and all replicates of the wild-type microglia (WT, n = 3), with a batch correction to account for separate differentiations. For a detailed account of biological and technical replicates see [Figure S4B](#). Using edgeR, we first performed TMM normalization of libraries. Significant differences were then determined using the genewise negative binomial generalized linear model (glmQLFit). Expression levels were considered statistically significant if the FDR value was  $\leq 0.05$ . EdgeR estimates dispersion from replicates using the quantile-adjusted conditional maximum likelihood method (qCML). Details of the statistical analysis can be found in the [Figure 3](#) legend. The individual samples compared can be found in [Figure 3B](#).